IRE Transactions on INFORMATION THEORY

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization.

Volume IT-6 **SEPTEMBER, 1960** **Number 4**

Published Quarterly

In This Issue

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $4.00.

IRE TRANSACTIONS® on INFORMATION THEORY is published in March, June, September, and December, by the IRE for the Professional Group on Information Theory, at 1 East 79th Street, New York 21, N. Y. In addition to these regular quarterly issues, Special Issues appear from time to time. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE-PGIT members, $2.30; IRE members, $3.45; nonmembers, $6.90.

INFORMATION THEORY

IRE Transactions on Information Theory

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization

Volume IT-6 | September, 1960 | Number 4

Published Quarterly

TABLE OF CONTENTS

On the Statistical Theory of Optimum Demodulation*

J. B. THOMAS†, MEMBER, IRE, AND E. WONG†, MEMBER, IRE

Summary—**The multidimensional demodulation problem is considered from the point of view of statistical estimation theory and *a posteriori* most probable signal estimates are derived. Correlated signals and noises are treated. This formulation yields a set of two matrix integral equations which must be solved for the optimum estimates.**

For amplitude modulation, the problem reduces to that of finding a set of time varying filters which are, again, solutions to a matrix integral equation. Special cases such as two-receiver systems, quadrature modulation, and single-sideband have particularly simple representations and are considered in some detail.

AN interesting problem in statistical communication theory is the "optimum" estimation of modulated intelligence in the presence of additive noise. For linear forms of modulation, the problem is essentially that of linear nonstationary filtering, and application of the minimum mean-squared error criterion leads to a reasonably simple integral equation.[1,2] Similarly, for nonlinear modulations, *e.g.*, FM, PM, etc., minimum mean-squared error nonlinear filtering theory can be applied.[3] However, even with simplifying restrictions,[4] the resulting mathematics is formidable and not usually amendable to explicit solutions. The methods of statistical estimation theory have been used to obtain *a posteriori most probable* estimates of generally modulated Gaussian signals in Gaussian noise.[5] This treatment results in two integral equations which specify the optimum receiver.

An extension of such estimation techniques to the multidimensional case is considered here. This extension treats the reception of more than one waveform, the estimation of more than one signal, and the case where signals and noises are correlated.

Formulation

The use of *a posteriori most probable* estimation is discussed in detail in the literature.[5-8] It suffices to state here that, given the received waveforms, those signals a chosen as estimates which have the greatest condition likelihood of occurrence.

Let the received waveforms be

$$\bar{r}(u) = \bar{m}[\bar{a}(u), u] + \bar{n}(u), \qquad t - T \leq u \leq t, \qquad (1$$

where $\bar{r}(u)$, $\bar{m}[\bar{a}(u), u]$, $\bar{a}(u)$, and $\bar{n}(u)$ are column vecto *e.g.*,

$$\bar{r}(u) = \begin{Bmatrix} r_1(u) \\ r_2(u) \\ \cdot \\ \cdot \\ \cdot \\ r_q(u) \end{Bmatrix}. \qquad (2$$

Here, $\bar{a}(u)$ represents the modulating signals and $\bar{n}(u$ additive noises. The components of both $\bar{a}(u)$ and $\bar{n}($ are assumed to be correlated Gaussian time series wi zero means. The vector $\bar{m}[\bar{a}(u), u]$ is a general modulatio function whose form depends on the modulating schem It is assumed that this modulation function is differentiab with respect to the elements $a_i(u)$.

In general, the noise vector $\bar{n}(u)$ will have q component as will $\bar{m}[\bar{a}(u), u]$. The signal vector $\bar{a}(u)$ will be taken have k components where k and q are not necessari related.

The problem is to find the set of $a_i(u)$, denoted $a_i^*(u)$, such that the conditional probability $p(\bar{a}/\bar{r})$ is maximum. Let the joint probability $p(\bar{a}, \bar{n}, \bar{r})$ be writt

$$p(\bar{a}, \bar{n}, \bar{r}) = p[(\bar{a}, \bar{n})/\bar{r}]p(\bar{r}) = p[\bar{r}/(\bar{a}, \bar{n})]p(\bar{a}, \bar{n}) \qquad (3$$

where $p(\bar{a}, \bar{n}, \bar{r})$ is the probability of the simultaneo realizations of $\bar{r}(u)$, $\bar{a}(u)$ and $\bar{n}(u)$ in the interval $t - T \leq u \leq t$, and a similar definition holds for the oth terms. Eq. (3) may be rewritten

$$p[(\bar{a}, \bar{n})/\bar{r}] = \frac{p[\bar{r}/(\bar{a}, \bar{n})]p(\bar{a}, \bar{n})}{p(\bar{r})}. \qquad (4$$

If it is noted that

$$p[(\bar{r}/(\bar{a}, \bar{n})] = \delta[\bar{n} - (\bar{r} - \bar{m})], \qquad (5$$

where $\delta(x)$ is the Dirac delta-function, then

$$p(\bar{a}/\bar{r}) = \frac{p[\bar{a}, (\bar{r} - \bar{m})]}{p(\bar{r})}. \qquad (6$$

Eq. (6) was obtained by integrating both sides of (4) wi respect to $\bar{n}$ and using the relationship of (1). For a giv set of received waveforms, $p(\bar{r})$ is a constant; therefore,

$$p(\bar{a}/\bar{r}) = k_1 p[\bar{a}, (\bar{r} - \bar{m})].$$

* Received by the PGIT, August 6, 1959.

† Dept. of Elec. Engrg., Princeton University, Princeton, N. J.

[1] R. C. Booton, Jr., "An optimization theory for time-varying linear systems with nonstationary statistical inputs," PROC. IRE, vol. 40, pp. 977–981; August, 1952.

[2] R. C. Booton, Jr., and M. H. Goldstein, Jr., "The design and optimization of synchronous demodulators," 1957 IRE WESCON CONVENTION RECORD, pt. 2, pp. 154–170.

[3] L. A. Zadeh, "Optimum nonlinear filters," *J. Appl. Phys.*, vol. 24, pp. 396–404; April, 1953.

[4] Such as restricting the nonlinear filter to be a one-convolution filter.

[5] D. C. Youla, "The use of maximum likelihood in estimating continuously modulated intelligence which has been corrupted by noise," IRE TRANS. ON INFORMATION THEORY, vol. IT-3, pp. 90–105; March, 1954.

[6] P. M. Woodward and I. L. Davies, "A theory of radar information," *Phil. Mag.*, ser. 7, vol. 41, pp. 1001–1017; October, 1950.

[7] P. M. Woodward and I. L. Davies, "Information theory and inverse probability in telecommunication," *Proc. IEE*, vol. 99, pp. 37–44; March, 1952.

[8] F. W. Lehan and R. J. Parks, "Optimum demodulation," 1953 IRE NATIONAL CONVENTION RECORD, pt. 8, pp. 101–103.

is desired to maximize this expression with respect to ne elements of $\bar{a}$.

Analysis

Define $\bar{x}(u)$ to be a column vector

$$\bar{x}(u) = \begin{Bmatrix} \bar{a}(u) \\ \bar{n}(u) \end{Bmatrix} \tag{8}$$

nd the associated covariance function matrix $\mathbf{R}(u, v)$ ith elements

$$R_{ij}(u, v) \triangleq E\{x_i(u)x_j(v)\} \tag{9}$$

here $E\{\ \}$ indicates the expectation of the bracketted uantity. It is apparent that $\mathbf{R}(u, v)$ can be partitioned as

$$\mathbf{R}(u, v) = \begin{Bmatrix} \mathbf{R}_{aa}(u, v) & \mathbf{R}_{an}(u, v) \\ \mathbf{R}_{na}(u, v) & \mathbf{R}_{nn}(u, v) \end{Bmatrix}. \tag{10}$$

It is convenient to use a multidimensional expansion ecently introduced,[9] and to write

$$\bar{x}(u) = \sum_{p=1}^{\infty} \alpha_p \bar{\varphi}_p(u), \qquad t - T \leq u \leq t, \tag{11}$$

here $\bar{\varphi}_p(u)$ is a column vector of $q + k$ components,

$$\bar{\varphi}_p(u) = \begin{Bmatrix} \varphi_p^{(1)}(u) \\ \varphi_p^{(2)}(u) \\ \vdots \\ \varphi_p^{(q+k)}(u) \end{Bmatrix}. \tag{12}$$

the $\bar{\varphi}_p(u)$ are the vector eigenfunctions of the matrix ntegral equation

$$(u) = \lambda \int_{t-T}^{t} \mathbf{R}(u, v)\bar{\varphi}(v)\, dv, \qquad t - T \leq u \leq t, \tag{13}$$

hen it can be shown[9] that these vectors $\bar{\varphi}_p(u)$ are orthog- nal in the sense that, after normalization

$$\int_{t-T}^{t} \bar{\varphi}_p(u) \cdot \bar{\varphi}_s(u)\, du = \delta_{ps}, \tag{14}$$

nd that the coefficients α_p are uncorrelated, *i.e.*,

$$E\{\alpha_p \alpha_s\} = \frac{1}{\lambda_p}\, \delta_{ps}. \tag{15}$$

Since the conditional probability $p(\bar{a}/\bar{r})$ is proportional the joint probability of the components of $\bar{x}(u)$,

$$p(\bar{a}/\bar{r}) \sim \exp\left(-\tfrac{1}{2} \sum_{p=1}^{\infty} \lambda_p \alpha_p^2\right). \tag{16}$$

herefore, in order to maximize $p(\bar{a}/\bar{r})$, it is sufficient to inimize the quantity $\sum_{p=1}^{\infty} \lambda_p \alpha_p^2$. It is shown in the ppendix that

$$\sum_{p=1}^{\infty} \lambda_p \alpha_p^2 = \int_{t-T}^{t} \int_{t-T}^{t} \bar{x}(u) \cdot \mathbf{Q}(u, v)\bar{x}(v)\, du\, dv, \tag{17}$$

[9] This expansion has been used by L. A. Zadeh and one of us in nnection with other work not yet published.

where the matrix $\mathbf{Q}(u, v)$ satisfies the integral equation

$$\int_{t-T}^{t} \mathbf{R}(u, v)\mathbf{Q}(v, w)\, dv = \delta(u - w)1, \qquad t - T \leq u, w \leq t, \tag{18}$$

1 being a unit matrix. It is apparent from (83) that the matrix $\mathbf{Q}(u, v)$ may be partitioned as

$$\mathbf{Q}(u, v) = \begin{bmatrix} \mathbf{Q}_{aa}(u, v) & \mathbf{Q}_{an}(u, v) \\ \mathbf{Q}_{na}(u, v) & \mathbf{Q}_{nn}(u, v) \end{bmatrix}. \tag{19}$$

It is now easy to minimize (16) with respect to the $a_i(u)$. By the familiar techniques of the calculus of variation, the following is obtained:

$$\int_{t-T}^{t} [\mathbf{Q}_{aa}(u, v) - \mathbf{M}(\bar{a}^*, u)\mathbf{Q}_{na}(u, v)]\bar{a}^*(v)\, dv = \int_{t-T}^{t} [\mathbf{M}(\bar{a}^*, u)\mathbf{Q}_{nn}(u, v) - \mathbf{Q}_{an}(u, v)] \cdot [\bar{r}(v) - \bar{m}(\bar{a}^*, v)]\, dv, \qquad t - T \leq u \leq t, \tag{20}$$

where the modulation matrix $\mathbf{M}(\bar{a}^*, u)$ has the elements

$$M_{ij}(\bar{a}^*, u) = \left.\frac{\partial m_j(\bar{a}, u)}{\partial a_i(u)}\right|_{\bar{a}=\bar{a}^*}. \tag{21}$$

Eq. (20) together with (18) is sufficient for the solution of the $a_i^*(u)$ in terms of the received waveforms $\bar{r}(u)$.

In the special case where the noises are uncorrelated with the signals, (21) and (18) can be used to obtain

$$\bar{a}^*(u) = \int_{t-T}^{t} \mathbf{R}_{aa}(u, v)\mathbf{M}(\bar{a}^*, v)\bar{g}(v)\, dv, \quad t - T \leq u \leq t, \tag{22}$$

and

$$\bar{r}(u) - \bar{m}(\bar{a}^*, u) = \int_{t-T}^{t} \mathbf{R}_{nn}(u, v)\bar{g}(v)\, dv, \qquad t - T \leq u \leq t, \tag{23}$$

where $\bar{g}(v)$ has been written for the expression

$$\int_{t-T}^{t} \mathbf{Q}_{nn}(v, w)[\bar{r}(w) - \bar{m}(\bar{a}^*, w)]\, dw.$$

In the one-dimensional case, (22) and (23) reduce to those obtained by Youla.[5]

In principle, the *a posteriori most probable* demodulator has been found. It is only necessary to specify the form of modulation and the covariance functions of the signals and noises. In practice, the solutions to the equations may be prohibitively difficult depending on the form of modulation.

Amplitude Modulation

General forms of amplitude modulation produce relatively simple expressions for the specifying equations and will be considered in some detail. In these cases, the modulation matrix $\mathbf{M}$ is not a function of the signals $a_i(t)$ and can be written as $\mathbf{M}(u)$. Then, the received waveforms are

$$\bar{r}(u) = \tilde{\mathbf{M}}(u)\bar{a}(u) + \bar{n}(u), \qquad t - T \leq u \leq t, \tag{24}$$

where $\tilde{\mathbf{M}}$ is the transpose of $\mathbf{M}$.

If, furthermore, the noises are uncorrelated with the signals, manipulation of (22) and (23) yields

$$\bar{a}^*(u) = \int_{t-T}^{t} \mathbf{W}(u, v)\bar{r}(v)\, dv, \qquad t - T \leq u \leq t, \tag{25}$$

and

$$\int_{t-T}^{t} \mathbf{W}(u, v)[\tilde{\mathbf{M}}(v)\mathbf{R}_a(v, w)\mathbf{M}(w) + \mathbf{R}_n(v, w)]\, dv = \mathbf{R}_a(u, w)\mathbf{M}(w), \qquad t - T \leq u, w \leq t, \tag{26}$$

where $\mathbf{W}(u, v)$ is a weighting function matrix determined from (26). Eq. (26) can also be derived as the specifying equation for the minimum mean-squared error nonstationary filter.

Eqs. (25) and (26) may be used to investigate a number of special cases of interest.

Case 1—Multireceiver Systems

When the noise level at the receiver itself is large compared to that of the transmission link, it is advantageous to consider multireceiver systems. Their advantage lies in the fact that the noises in the various inputs are uncorrelated, while the signals are either highly correlated or the same. Applications for these types of systems occur, for example, in the field of radio astronomy.[10,11]

Let us consider a two-receiver system where the received waveforms are

$$r_1(u) = M(u)a(u) + n_1(u), \qquad t - T \leq u \leq t \tag{27}$$

and

$$r_2(u) = M(u)a(u) + n_2(u), \qquad t - T \leq u \leq t. \tag{28}$$

Then (25) and (26) become

$$a^*(u) = \int_{t-T}^{t} W_1(u, v)r_1(v)\, dv + \int_{t-T}^{t} W_2(u, v)r_2(v)\, dv, \tag{29}$$

and

$$\int_{t-T}^{t} W_1(u, v)[M(v)M(w)R_a(v, w) + R_{n11}(v, w)]\, dv + \int_{t-T}^{t} W_2(u, v)M(v)M(w)R_a(v, w)\, dv = R_a(u, w)M(w). \tag{30}$$

$$\int_{t-T}^{t} W_2(u, v)[M(v)M(w)R_a(v, w) + R_{n22}(v, w)]\, dv + \int_{t-T}^{t} W_1(u, v)M(v)M(w)R_a(v, w)\, dv = R_a(u, w)M(w). \tag{31}$$

[10] R. H. Dicke, "The measurement of thermal radiation at microwave frequencies," *Rev. Sci. Instr.*, vol. 17, pp. 268–275; July, 1946.

[11] S. J. Goldstein, "A comparison of two radiometer circuits," PROC. IRE, vol. 43, pp. 1663–1666; November, 1955.

It is interesting to note that if $R_{n11} = R_{n22} \triangleq R_n$, then th symmetry of (30) and (31) implies that

$$W_1(u, v) = W_2(u, v) \triangleq W(u, v), \tag{3}$$

and (30) and (31) reduce to

$$\int_{t-T}^{t} W(u, v)[2M(v)M(w)R_a(v, w) + R_n(v, w)]\, dv = R_a(u, w)M(w), \tag{3}$$

while the corresponding equation for one dimension is

$$\int_{t-T}^{t} W(u, v)[M(v)M(w)R_a(v, w) + R_n(v, w)]\, dv = R_a(u, w)M(w). \tag{3}$$

A compasison of (33) and (34) indicates the advantag of a two-receiver system. Effectively, the signal lev relative to noise is doubled.

Case 2—Multiplex Systems

Various multiplex modulation schemes are used i communication. They have the common characterist that more than one signal is transmitted simultaneousl on a time or frequency sharing basis.[12]

Quadrature Modulation

One of the most familiar examples of multiplexing quadrature modulation, which, in the formulation di cussed here, has a particularly simple representation.

Let the received waveform be

$$r(u) = \cos \omega_0 u a_1(u) + \sin \omega_0 u a_2(u) + n(u), t - T \leq u \leq (3$$

Then, (25) and (26) become

$$a_1^*(u) = \int_{t-T}^{t} W_1(u, v)r(v)\, dv, \tag{3}$$

$$a_2^*(u) = \int_{t-T}^{t} W_2(u, v)r(v)\, dv, \tag{3}$$

and

$$\int_{t-T}^{t} W_1(u, v)[\cos \omega_0 v R_{a11}(v, w) \cos \omega_0 w + \cos \omega_0 v R_{a12}(v, w) \sin \omega_0 w + \sin \omega_0 v R_{a21}(v, w) \cos \omega + \sin \omega_0 v R_{a22}(v, w) \sin \omega_0 w + R_n(v, w)]\, dv = R_{a11}(u, w) \cos \omega_0 w + R_{a12}(u, w) \sin \omega_0 w, \tag{3}$$

$$\int_{t-T}^{t} W_2(u, v)[\cos \omega_0 v R_{a11}(v, w) \cos \omega_0 w + \cos \omega_0 v R_{a12}(v, w) \sin \omega_0 w + \sin \omega_0 v R_{a21}(v, w) \cos \omega + \sin \omega_0 v R_{a22}(v, w) \sin \omega_0 w + R_n(v, w)]\, dv = R_{a21}(u, w) \cos \omega_0 w + R_{a22}(u, w) \sin \omega_0 w. \tag{3}$$

[12] H. S. Black, "Modulation Theory," D. Van Nostrand Inc., New York, N. Y.; 1953.

should be noted that the kernels of (39) and (40) are identical.

ingle Sideband

Although single-sideband amplitude modulation is asically a one-dimensional problem, it can be conveniently treated as a special case of quadrature modulation. In this case,

$$u) = \cos \omega_0 u a(u) + \sin \omega_0 u \hat{a}(u) + n(u),$$

$$t - T \leq u \leq t, \tag{41}$$

here $\hat{a}(u)$, the Hilbert transform of $a(u)$, is defined[13] by

$$\hat{a}(u) \triangleq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a(v)}{u - v}\, dv. \tag{42}$$

In the case when the signal and noise are stationary, the presentation simplifies even further. If we define $(u) \triangleq E\{a(t)\ a(t + u)\}$, then the following relationships re easily derived:

$$E\{a(t)\hat{a}(t + u)\} = \hat{R}(u), \tag{43}$$

$$E\{\hat{a}(t)a(t + u)\} = -\hat{R}(u), \quad \text{and} \tag{44}$$

$$E\{\hat{a}(t)\hat{a}(t + u)\} = R(u). \tag{45}$$

sing these expressions, we find that (39) reduces to

$$\int_{t-T}^{t} W_1(u, v)[R_a(w - v) \cos \omega_0(w - v) + \hat{R}_a(w - v) \sin \omega_0(w - v) + R_n(w - v)]\, dv = R_a(w - u) \cos \omega_0 w + \hat{R}_a(w - u) \sin \omega_0 w. \tag{46}$$

n this case, $W_2(u, v)$ is of no interest since it gives the stimate of $\hat{a}(u)$. It should be noted that $\hat{R}_a(u)$ is an odd unction of u, and therefore, the kernel of (46) remains ymmetric with respect to the variables u and v, as it ust.

Quadrature Modulation Example

In the special case where $a_1(u)$, $a_2(u)$ and $n(u)$ are ationary, and where $a_1(u)$ and $a_2(u)$ are uncorrelated nd have the same autocorrelation function, (39) and 0) reduce to

$$\int_{t-T}^{t} W_1(u, v)[R_a(w - v) \cos \omega_0(w - v) + R_n(w - v)]\, dv = R_a(w - u) \cos \omega_0 w \tag{47a}$$

nd

$$\int_{t-T}^{t} W_2(u, v)[R_a(w - v) \cos \omega_0(w - v) + R_n(w - v)]\, dv = R_a(w - u) \sin \omega_0 w, \tag{47b}$$

here

$$R_a(w - v) = R_{a_{11}}(v, w) = R_{a_{22}}(v, w). \tag{48}$$

[13] E. C. Titchmarch, "Introduction to the Theory of Fourier tegrals," Oxford University Press, London, England; 1937.

Note that the integral equations (47a) and (47b) have kernels which are functions of the difference of two variables and can be solved easily by standard techniques.[14]

In order to obtain an indication of the forms of the optimum demodulators, we shall give an example with explicit solutions. Let

$$R_a(u) = A_0 \frac{\alpha}{2} e^{-\alpha|u|}, \tag{49}$$

$$R_n(u) = N_0\, \delta(u), \tag{50}$$

and the received waveform be given for the range $-\infty$ to t.

Then, with a change in variables, (37) and (47a) become

$$a_1^*(t) = \int_0^{\infty} W_1'(t, u) r(t - u)\, du \tag{51}$$

and

$$\int_0^{\infty} W_1'(t, u)\left[\frac{A_0\alpha}{2} e^{-\alpha|v-u|} \cos \omega_0(v - u) + N_0\, \delta(v - u)\right] du = \frac{A_0\alpha}{2} e^{-\alpha v} \cos \omega_0(t - v). \tag{52}$$

Eq. (52) can be solved in the usual way[14] and $W_1'(t, u)$ is found to have the form

$$W_1'(t, u) = h_{11}(u) \cos \omega_0(t - u) + h_{12}(u) \sin \omega_0(t - u). \tag{53}$$

In other words, the receiver can be represented as shown in Fig. 1. This is a form of synchronous receiver with

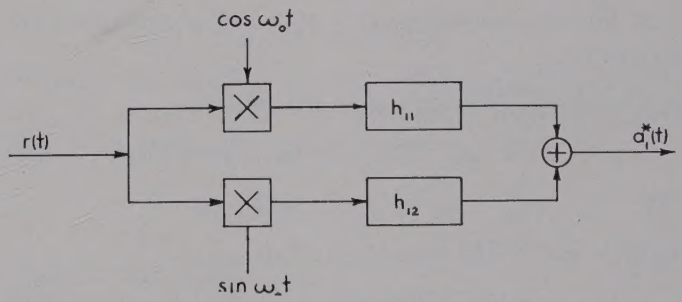


Fig. 1—An optimum quadrature demodulator.

specific stationary filters. For convenience, we define the constants

$$\epsilon \triangleq \frac{A_0}{2N_0} \tag{54}$$

and

$$\beta \triangleq \frac{\alpha}{\omega_0}, \tag{55}$$

and consider two cases.

[14] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645–655; July, 1950.

1) For the case where $\beta^2 \leq 4(1 + \epsilon)/\epsilon^2$, the filters are given by

$$h_{11}(u) = \omega_0 e^{-a\alpha u}[K_1 \cos \omega_0(1 - b)u - K_2 \cos \omega_0(1 + b)u + K_3 \sin \omega_0(1 - b)u - K_4 \sin \omega_0(1 + b)u] \quad (56)$$

and

$$h_{12}(u) = \omega_0 e^{-a\alpha u}[K_3 \cos \omega_0(1 - b)u - K_4 \cos \omega_0(1 + b)u - K_1 \sin \omega_0(1 - b)u + K_2 \sin \omega_0(1 + b)u], \quad (57)$$

with constants

$$K_1 = \frac{\beta(a - 1)[\beta^2(a - 1)^2 + (1 + b)^2]}{2b}, \quad (58)$$

$$K_2 = \frac{\beta(a - 1)[\beta^2(a - 1)^2 + (1 - b)^2]}{2b}, \quad (59)$$

$$K_3 = \frac{(1 - b)[\beta^2(a - 1)^2 + (1 - b)^2]}{2b}, \quad (60)$$

$$K_4 = \frac{(1 + b)[\beta^2(a - 1)^2 + (1 - b)^2]}{2b}, \quad (61)$$

and

$$a = \frac{1}{\beta}\{\tfrac{1}{2}[(\beta^2 + 1)^2 + 2\epsilon\beta^2(\beta^2 + 1)]^{1/2} + (1 + \epsilon)\beta^2 - 1\}^{1/2} \quad (62)$$

and

$$b = \{\tfrac{1}{2}[(\beta^2 + 1)^2 + 2\epsilon\beta^2(\beta^2 + 1)]^{1/2} - (1 + \epsilon)\beta^2 + 1\}^{1/2}. \quad (63)$$

2) For the case where $\beta^2 > 4(1 + \epsilon)/\epsilon^2$, the filters are given by

$$h_{11}(u) = \omega_0 e^{-a'\alpha u}(K_2' \sin \omega_0 u - K_1' \cos \omega_0 u) + \omega_0 e^{-b'\alpha u}(K_3' \cos \omega_0 u - K_4' \sin \omega_0 u) \quad (64)$$

and

$$h_{12}(u) = \omega_0 e^{-a'\alpha u}(K_2' \cos \omega_0 u + K_1' \sin \omega_0 u) - \omega_0 e^{-b'\alpha u}(K_3' \sin \omega_0 u + K_4' \cos \omega_0 u), \quad (65)$$

where

$$a' = \frac{1}{\beta}\{(1 + \epsilon)\beta^2 - 1 - \beta[\beta^2\epsilon^2 - 4(1 + \epsilon)]^{1/2}\}^{1/2} \quad (66)$$

$$b' = \frac{1}{\beta}\{(1 + \epsilon)\beta^2 - 1 + \beta[\beta^2\epsilon^2 - 4(1 + \epsilon)]^{1/2}\}^{1/2} \quad (67)$$

and

$$K_1' = \frac{\beta^2(a' - 1)^2 + 1}{\beta(b' - a')}, \quad (68)$$

$$K_2' = \frac{(b' - 1)}{(b' - a')}[\beta^2(a' - 1)^2 + 1], \quad (69)$$

$$K_3' = \frac{\beta^2(b' - 1)^2 + 1}{\beta(b' - a')}, \quad \text{and} \quad (70)$$

$$K_4' = \frac{a' - 1}{b' - a'}[\beta^2(b' - 1)^2 + 1]. \quad (7$$

It should be noted that, despite their complexity, th filters can be synthesized as lumped-constant R-L- networks for any given β and ϵ.

It is interesting to consider some limiting cases of th example:

1) $\epsilon \to 0$ (very small signal power),

$$h_{11}(u) \to \omega_0\beta\epsilon e^{-\alpha u}, \quad (7$$

$$h_{12}(u) \to 0 \quad (7$$

to the first order in ϵ.

2) $\epsilon \to \infty$ (noise power becomes negligible),

$$h_{11}(u) \to \frac{(k - \alpha)^2 + \omega_0^2}{\omega_0} \sin \omega_0 u e^{-ku} + \delta(u), \quad (7$$

where

$$k = (\alpha^2 + \omega_0^2)^{1/2} \quad (7$$

and

$$h_{12}(u) \to \frac{2k(k - \alpha)}{\omega_0} \cos \omega_0 u e^{-ku} + \frac{k - \alpha}{\omega_0}\delta(u). \quad (7$$

In the same way as before, the optimum estimat $a_2^*(t)$ can be found to be

$$a_2^*(t) = \int_0^\infty W_2'(t, u)r(t - u)\, du \quad (7$$

with

$$W_2'(t, u) = h_{21}(u) \cos \omega_0(t - u) + h_{22}(u) \sin \omega_0(t - u). \quad (7$$

It should be noted that $h_{21}(u)$ and $h_{22}(u)$ are simpl related to $h_{12}(u)$ and $h_{11}(u)$; in fact,

$$h_{21}(u) = -h_{12}(u), \quad (7$$

and

$$h_{22}(u) = h_{11}(u). \quad (8$$

AM Demodulation with Delay

With present techniques, most of the integral equatio involved in optimum demodulation are difficult to solv explicitly. However, if a reasonable delay can be tolerate approximate solutions to a large class of AM proble can be obtained. The demodulator is found to be a sy cronous demodulator followed by a type of Wiener filte Problems of this nature have been treated in some detail for the one-dimensional case. Extensions to multi-c mensional cases are straightforward.[16]

[15] J. B. Thomas, "On the Statistical Design of Demodulati Systems for Signals in Additive Noise," Stanford University Ele tronics Res. Lab., Stanford, Calif., Tech. Rept. No. 88; August, 19

[16] J. B. Thomas, T. R. Williams, J. Wolf and E. Wong, "T demodulation of AM signals in noise," *Proc. 1959 IRE Conventi on Military Electronics*, pp. 138–146.

Nonlinear Modulations

For nonlinear forms of modulation such as FM and PM, the problem of optimum estimation cannot be reduced to that of finding a time varying filter. In general, one has to consider the solution of (20) for $a^*(u)$. Although this equation is not usually amenable to explicit solution, it is in a form that can be treated by analog techniques. Indeed, if feedback is allowed in the system, it essentially specifies the demodulator. Some work along this line has been initiated.[17]

Problems in Carrier Specification

In this formulation, the phases, amplitudes and frequencies of the carriers are assumed known. In practice, this knowledge must be obtained frequently either by operating on the received waveforms or by transmitting the carriers over a separate channel. Both of these methods involve errors due to noise and thus cause additional errors in the estimation of signals. Such difficulties are common to all synchronous receiver systems.

Appendix

With the use of the orthonormality condition given by (14), the coefficients of expansion α_p can be expressed as

$$\alpha_p = \int_{t-T}^{t} \bar{\varphi}_p(u) \cdot \bar{x}(u)\, du. \tag{81}$$

Therefore, the sum $\sum_{p=1}^{\infty} \lambda_p \alpha_p^2$ is evaluated to be

$$\sum_{p=1}^{\infty} \lambda_p \alpha_p^2 = \sum_{p=1}^{\infty} \lambda_p \int_{t-T}^{t} \int_{t-T}^{t} \left(\sum_{i=1}^{N} \varphi_p^{(i)}(u) x_i(u) \cdot \sum_{j=1}^{N} \varphi_p^{(j)}(v) x_j(v) \right) du\, dv \tag{82}$$

where $N = q + k$. Now, define the matrix $\mathbf{Q}(u, v)$ by the relationship

$$Q_{ij}(u, v) = \sum_{p=1}^{\infty} \lambda_p \varphi_p^{(i)}(u) \varphi_p^{(j)}(v). \tag{83}$$

Then, the sum $\sum_{p=1}^{\infty} \lambda_p \alpha_p^2$ becomes

$$\sum_{p=1}^{\infty} \lambda_p \alpha_p^2 = \int_{t-T}^{t} \int_{t-T}^{t} \bar{x}(u) \cdot \mathbf{Q}(u, v) \bar{x}(v)\, du\, dv. \tag{84}$$

The matrix $\mathbf{Q}(u, v)$ is related to the covariance function matrix $\mathbf{R}(u, v)$. This relationship becomes clear when the integral

$$\sum_{j=1}^{N} \int_{t-T}^{t} R_{ij}(u, v) Q_{jk}(v, w)\, dv$$

is examined. With the substitution of (83) for $Q_{jk}(v, w)$ and the use of (13), this integral becomes

$$\begin{aligned} \sum_{j=1}^{N} \int_{t-T}^{t} R_{ij}(u, v) Q_{jk}(v, w)\, dv &= \sum_{p=1}^{\infty} \lambda_p \varphi_p^{(k)}(w) \int_{t-T}^{t} \sum_{j=1}^{N} R_{ij}(u, v) \varphi_p^{(j)}(v)\, dv \\ &= \sum_{p=1}^{\infty} \varphi_p^{(i)}(u) \varphi_p^{(k)}(w). \end{aligned} \tag{85}$$

The sum on the right-hand side above satisfies the identity

$$\sum_{p=1}^{\infty} \varphi_p^{(i)}(u) \varphi_p^{(k)}(w) \equiv \delta_{ik}\, \delta(u - w). \tag{86}$$

To prove this identity, multiply both sides of (86) by $\varphi_p^{(i)}(u)$, sum over the index i, and integrate with respect to u. With the use of the orthonormality condition, this procedure yields the identity

$$\varphi_p^{(k)}(w) \equiv \varphi_p^{(k)}(w),$$

showing the validity of (85). Therefore, the matrix $\mathbf{Q}(v, w)$ is related to the covariance function matrix $\mathbf{R}(u, v)$ by matrix integral equation

$$\int_{t-T}^{t} \mathbf{R}(u, v) \mathbf{Q}(v, w)\, dv = \delta(u - w)\mathbf{1}, \tag{87}$$

$\mathbf{1}$ being a unit matrix.

---

[17] R. Jaffe and E. Rechtin, "Design and performance of phase locked circuits capable of near optimum performance over a wide range of input signal and noise levels," IRE Trans. on Information Theory, vol. IT-1, pp. 66–76; March, 1955.

On the Mean-Square Noise Power of an Optimum Continuous Filter for Correlated Noise*

MARVIN BLUM†, MEMBER, IRE

Summary—**This paper presents the equations for the mean-square error of the output of a continuous finite memory filter. The filter output error is unbiased for arbitrary input polynomials up to degree n, and has minimum variance. The input is taken as a polynomial of degree n plus random stationary noise. Noise processes are considered, 1) where the noise is exponentially correlated, and 2) in the white noise case. The solution for a desired output which is an arbitrary fixed linear operation on the input polynomial is given.**

Tables and graphs of the mean-square error for the derivative and prediction operator for the 0th, 1st, and 2nd derivatives are presented, and for input polynomials up to the 6th degree.

Introduction

In a paper by Zadeh and Ragazzini,[1] a solution for the optimum continuous filter in a minimum variance sense was given. In this paper, the authors developed, as an illustrative example, the detailed equations for the mean-square error output of a first order polynomial passing filter for exponentially correlated noise input. However, an attempt to extend the details of the solution to higher order filters and maintain an analytic solution leads to prohibitive labors because of the necessity of inverting high order matrices whose elements are each functions. To circumvent this difficulty, a method is used which is based on a generalization of the optimum filter as presented in a previous paper.[2] In this solution a set of orthogonal polynomials is defined such that the submatrices, depending upon the order of the filter, are diagonalized and an analytic solution becomes feasible for any order of the filter.

White Noise Solution

Let the input to a continuous filter with finite memory over the interval $(0, T)$ be

$$S(t) = P(t) + N(t) \tag{1}$$

where

$$P(t) = \sum_{k=0}^{n} a_k P_k(t) \qquad -\infty \leq t \leq +\infty, \tag{2}$$

and $P_k(t)$ is a modified Legendre polynomial[3] given by

$$P_k(t) \equiv \sum_{j=0}^{k} (-1)^j \binom{k}{j}\binom{k+j}{j}\left(\frac{t}{T}\right)^j. \tag{3}$$

Then it can be shown that these polynomials are orthogonal with respect to integration over the interval $(0, T)$ *e.g.*,

$$\int_0^T P_k(t)P_h(t)\,dt = 0 \qquad k \neq h \tag{4}$$

$$\int_0^T [P_k(t)]^2\,dt = \frac{T}{2k+1} \equiv S_k. \tag{5}$$

It is easily shown that the right hand side is unchanged if $(T - t)$ is substituted for t in the left hand side of (4) and (5). Let $N(t)$ be a white noise process with ensemble average equal to zero for all t, such that the autocorrelation function is defined by $\delta(\tau)$, and the spectral density function is unity.

Using the notation of Blum's previous paper,[2] let

$$Q_k \equiv \int_{-\infty}^{+\infty} k(t)P_k(T - t)\,dt \qquad k = 0, 1, \cdots, n \tag{6}$$

where the unspecified desired linear transformation on the input data defines $k(t)$. Note that by Blum's equation (14)[4]

$$Q_k \equiv \int_0^T P_k(T - x)W(x)\,dx \qquad k = 0, 1, \cdots, n, \tag{7}$$

where $W(x)$ is the weighting function of the optimum filter. Let

$$S_{k,l} = \int_0^T P_k(T - t)P_l(T - t)\,dt \qquad k, l = 0, 1, \cdots, n. \tag{8}$$

Then by (4), the $(n + 1) \times (n + 1)$ matrix S whose elements are $S_{k,l}$ is given by a diagonal matrix whose elements are $S_0, S_1, \cdots, S_n$ [see (5)]. The rms output error by (42)[4] becomes

$$\sigma_\infty^2 = \sum_{k=0}^{n} \frac{Q_k^2}{S_k}. \tag{9}$$

* Received by the PGIT, April 8, 1959.

† System Development Corp., Santa Monica, Calif.

[1] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645–655; July, 1950.

[2] Marvin Blum, "Generalization of the class of nonrandom inputs of the Zadeh-Ragazzini prediction model," IRE TRANS. ON INFORMATION THEORY, vol. IT-2, pp. 76–81; June, 1956.

[3] The modification consists of substituting $y = t/T$ where $0 \leq y \leq 1$.

[4] Equation numbers of author's paper, footnote 2.

he weighting function is given by (41),[4] and becomes

$$W(x) = \sum_{k=0}^{n} \frac{P_k(T-x)}{S_k} Q_k \qquad 0 \le x \le T. \tag{10}$$

n particular, when the desired output is the Lth derivative f the input at $T - \alpha$, then

$$Q_k = \frac{d^L}{dt^L} P_k(T-t)\Big|_{t=\alpha} \equiv P_k^{(L)}(T-\alpha), \tag{11}$$

o that the rms ouput error (9) is given by

$$\sigma_\infty^2 = \sum_{k=L}^{n} \frac{[P_k^{(L)}(T-\alpha)]^2}{S_k}. \tag{12}$$

et

$$y = \frac{T-\alpha}{T}, \qquad t = T - \alpha;$$

hen

$$P_k^{(L)}(T-\alpha) \equiv \frac{1}{T^L}\frac{d^L}{dy^L} P_k(y) = \frac{1}{T^L} P_k^{(L)}(y), \tag{13}$$

o that (12) becomes

$$T^{(2L+1)}\sigma_\infty^2 = \sum_{k=L}^{n} [P_k^{(L)}(y)]^2(2k+1) \equiv H_{n,L}^2(y). \tag{14}$$

Properties of $H_{n,L}^2(y)$, the Mean-Square Proportionality Factor

The prediction parameter α has the following interpretation. When $\alpha = 0$, $y = 1$, and the estimate corresponds to an end point or zero lag smoothing. When $= T/2$, $y = 1/2$, the output of the filter at time T is n estimate of the midpoint of the data interval. When < 0 so that $y > 1$, the output of the filter is an estimate f a predicted value of the input. The function $H_{n,L}^2(y)$ atisfies the following relationships:

$$H_{n+1,L}^2(y) \ge H_{n,L}^2(y), \tag{15}$$

$$H_{n,L}^2(1-y) = H_{n,L}^2(y), \quad \text{and} \tag{16}$$

$$H_{n,L}^2(y) = \sum_{k=0}^{n} a_{2k}(y - \tfrac{1}{2})^{2k}. \tag{17}$$

n (15) the equality sign holds only for the roots of $P_k(y)$ nd certain adjacent pairs of integers. Eqs. (16) and (17) tate that $H_{n,L}^2(y)$ is symmetric about $y = 1/2$ and is epresentable as a polynomial of degree $2n$ in y. As a onsequence of (17), $H_{n,L}^2(y)$ is a polynomial of degree $2n$ n y but only of degree n in $(y - 1/2)^2$. Thus for purposes f interpolation it is more convenient to use a divided ifference interpolation formula using the variable

$$z = (y - \tfrac{1}{2})^2. \tag{18}$$

The orthogonal polynomials $P_k(y)$, $k = 0, 1, \cdots, 8$ re listed in Milne.[5] A listing of the divided difference interpolation (and extrapolation) formulas are found in Appendix I. Graphs of $H_{n,l}(y)$ vs. y are plotted for $n = 0, 1, \cdots, 6$, $l = 0, 1, 2$ and $1/2 \le y \le 7/2$.

Exponentially Correlated Noise

The solution for the mean-square error of estimation will be obtained for the case when the autocorrelation function is given in two forms:

a) σ_a^2 corresponding to the correlation function $\psi_a(\tau) = [a/2]\, e^{-a|\tau|}$, and

b) θ^2 corresponding to the correlation function $\theta(\tau) = e^{-a|\tau|}$. (19)

The solution for form a) will approach the solution for the white noise case as $a \to \infty$. The details of the solution for form b) will be given.

In (23),[4] it is shown that the weighting function $W(x)$, defining the optimum unbiased filter with minimum variance at $T_0 = T$, must satisfy the integral equation

$$\int_0^T W(x)\theta(t-x)\,dx = \sum_{k=0}^{n} \lambda_k P_k(T-t)$$

$$0 \le t \le T. \tag{20}$$

The explicit solution of (20) is shown to be by (32)[4] and (35)[4] of the form (21), where a standard interval $(0, T)$ will be used without loss of generality, and, thus, the unit step function may be deleted. The coefficients μ_k are independent of time and are to be determined.

$$W(x) = \sum_{k=0}^{n} \mu_k P_k(T-x) + c\,\delta(x) + D\,\delta(T-x)$$

$$0 \le x \le T. \tag{21}$$

To determine the coefficients μ_k, C and D of (21) one uses the $n + 1$ constraint relationships of (14)[4]

$$\int_0^T P_k(T-x)W(x)\,dx = Q_k$$

$$k = 0, 1, 2, \cdots n. \tag{22}$$

One can substitute (21) into (20) to obtain two linear homogeneous equations involving the C and D. These come about because (20) is an identity in t, and in substituting into the left-hand side of (20), one generates functions $\theta(t)$ and $\theta(T-t)$. For the equation to remain an identity, the coefficients of these terms must be set equal to zero since these functions do not appear on the right hand side of (20).

Homogeneous Equations for C and D

On substituting (21) into (20) one obtains,

$$\left.\begin{aligned} \sum_{k=0}^{n} \mu_k \int_0^T \{P_k(T-x) + c\,\delta(x) + D\,\delta(T-x)\} \\ \cdot e^{-a|t-x|}\,dx \\ = \sum_{k=0}^{n} \lambda_k P_k(T-t). \end{aligned}\right\} \tag{23}$$

[5] W. E. Milne, "Numerical Calculus," Princeton University ress, Princeton, N. J., 3rd ed., p. 260; 1949.

Let us evaluate the kth component on the left-hand side of (23), then

$$\int_0^T P_k(T-x)e^{-a|t-x|}\,dx = -\theta(t)\sum_{L=0}^{k}\frac{b_{k,L}}{T^L}\sum_{i=0}^{L}\frac{T^{L-i}(L)^{(i)}}{a^{i+1}} + \sum_{L=0}^{k}\frac{b_{k,L}}{T^L}\sum_{i=0}^{L}\frac{(T-t)^{L-i}(L)^{(i)}}{a^{i+1}}\{1+(-1)^i\} - \theta(T-t)\sum_{L=0}^{k}\frac{b_{k,L}}{T^L}\frac{(-1)^L L!}{a^{L+1}} \tag{24}$$

where the substitutions

$$(m)^{(i)} = (m)(m-1)\cdots(m-j+1), \quad (m)^0 = 1$$

$$P_k(t) = \sum_{L=0}^{k} b_{k,L}\frac{t^L}{T^L} \qquad 0 \le t \le T \tag{25}$$

$$b_{k,L} = (-1)^L\binom{k}{L}\binom{k+L}{L}$$

have been made.

Performing the integration as indicated in (23) and setting the coefficients of $\theta(t)$ and $\theta(T-t)$ equal to zero, one obtains the two equations

$$\left.\begin{aligned} C + \sum_{k=0}^{n}\mu_k A_k &= 0 \\ D + \sum_{k=0}^{n}\mu_k B_k &= 0 \end{aligned}\right\} \tag{26}$$

where from (24),

$$A_k = -\frac{1}{a}\sum_{L=0}^{k}\sum_{i=0}^{L} b_{k,L}(aT)^{-i}(L)^{(i)} \tag{27}$$

$$B_k = -\frac{1}{a}\sum_{L=0}^{k} b_{k,L}(-aT)^{-L}L!. \tag{28}$$

Substituting the weighting sequence from (21) into (22), one obtains the system of linear equations

$$\mu_k S_k + CP_k(T) + DP_k(0) = Q_k \quad k = 0, 1, 2, \cdots n. \tag{29}$$

This system of linear equations can be solved by inverting a matrix by the method of submatrices.[6] Let the equations (26) and (29) be written as

$$\left|\begin{array}{cccc|c} \multicolumn{4}{c|}{|\alpha_{11}|} & |\alpha_{12}| \\ S_0 & 0 & 0 \cdots 0 & & P_0(T)P_0(0) \\ 0 & S_1 & 0 \cdots 0 & & P_1(T)P_1(0) \\ 0 & 0 & S_2 \cdots 0 & & P_2(T)P_2(0) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 \cdots S_n & & P_n(T)P_n(0) \\ \hline A_0 & A_1 & A_2 \cdots A_n & & 1 \quad 0 \\ B_0 & B_1 & B_2 \cdots B_n & & 0 \quad 1 \\ \multicolumn{4}{c|}{|\alpha_{21}|} & |\alpha_{22}| \end{array}\right| \left|\begin{array}{c} \mu_1 \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \\ \hline C \\ D \end{array}\right| = \left|\begin{array}{c} Q_0 \\ Q_1 \\ Q_2 \\ \vdots \\ Q_n \\ \hline 0 \\ 0 \end{array}\right|. \tag{30}$$

The notation as indicated by $\bar{\alpha}$ is an $(n+3)\times(n+3)$ matrix with submatrices as follows:

a) α_{11} is a diagonal $(n+1)\times(n+1)$ submatrix with elements S_i

b) α_{12} is an $(n+1)\times(2)$ submatrix,

c) α_{21} is a $(2)\times(n+1)$ submatrix, and

d) α_{22} is a $(2)\times(2)$ identity matrix.

Then $B = \bar{\alpha}^{-1}$ (the inverse matrix) is given by

$$B = \left|\begin{array}{c|c} B_{11}(n+1\times n+1) & B_{12}(n+1\times 2) \\ \hline B_{21}(2\times n+1) & B_{22}(2\times 2) \end{array}\right|. \tag{31}$$

Let

$$\left.\begin{aligned} X &= \alpha_{11}^{-1}\alpha_{12} \\ Y &= \alpha_{21}\alpha_{11}^{-1} \\ \bar{\theta} &= \alpha_{22} - Y\alpha_{12} \end{aligned}\right\}, \tag{32}$$

then

$$\left.\begin{aligned} B_{11} &= \alpha_{11}^{-1} + X\bar{\theta}^{-1}Y \\ B_{12} &= -X\bar{\theta}^{-1} \\ B_{21} &= -\bar{\theta}^{-1}Y \\ B_{22} &= \bar{\theta}^{-1} \end{aligned}\right\}, \tag{33}$$

where it is assumed that $\bar{\theta}$ and α_{11} are nonsingular. Using the relationships

$$\left.\begin{aligned} A_k &= (-1)^k B_k \\ P_k(0) &= 1 \\ P_k(T) &= (-1)^k \end{aligned}\right\} \qquad k = 0, 1, 2, \cdots n, \tag{34}$$

then

$$B_{11} = S^{-1} + H \tag{35}$$

where S^{-1} is the $(n+1)\times(n+1)$ diagonal matrix with elements S_i^{-1} and H is a $(n+1)\times(n+1)$ matrix with elements

$$H_{2k,2j} = \frac{2B_{2k}}{S_{2j}S_{2k}(1-2X_1^{(n)})}$$

[6] R. A. Frazer, W. J. Duncan, and A. R. Collar, "Elementary Matrices," Cambridge University Press, Cambridge, Eng., pp. 112–118; 1955.

r even columns and rows;

$$H_{2k+1,2j+1} = \frac{2B_{2k+1}}{S_{2j+1}S_{2k+1}(1 - 2X_2^{(n)})} \tag{36}$$

r odd columns and rows; and zero, for rows and columns hich are not jointly even or odd. The quantities $x_1^{(n)}$ nd $x_2^{(n)}$ are defined by

$$\left.\begin{aligned} x_1^{(n)} &= \sum_{k=0}^{[n/2]} \frac{B_{2k}}{S_{2k}} \qquad n = 0, 1, 2, \cdots \\ x_2^{(n)} &= \sum_{k=0}^{[(n-1)/2]} \frac{B_{2k+1}}{S_{2k+1}} \qquad n = 1, 2, \cdots \end{aligned}\right\}, \tag{37}$$

here [ ] is the largest integer in the bracket *e.g.*,

$$\left[\frac{5}{2}\right] = 2, \qquad \left[\frac{6}{2}\right] = 3.$$

olving (30) for μ_k, C and D, one has,

$$\left.\begin{aligned} C &= \frac{1}{[1 - 2X_2^{(n)}]} \sum_{k=0}^{n} \frac{Q_k B_k}{S_k} \\ D &= -C \\ \mu &= B_{11}Q \end{aligned}\right\} \tag{38}$$

here μ and Q are column matrices respectively of u_k and $_k$.

Evaluation of the Mean-Square Error for Exponentially Correlated Noise

To evaluate θ^2 for correlation form b) [θ^2 corresponding o the correlation function $\theta(\tau) = e^{-a|\tau|}$], one must etermine the relationships between the μ_k and the λ_k of 20), since it is easily shown that

$$\theta^2 = \sum_{k=0}^{n} Q_k \lambda_k. \tag{39}$$

Using the polynomial component of the right hand side f (24), one obtains

$$\sum_{=0}^{n} \sum_{L=0}^{k} \sum_{i=0}^{L} \frac{\mu_k b_{k,L}(T - t)^{L-i}(L)^{(i)}\{1 + (-1)^i\}}{T^L a^{i+1}} \equiv \sum_{k=0}^{n} \lambda_k P_k(T - t) \qquad 0 \le t \le T. \tag{40}$$

y suitably combining terms in the left hand side of (40), ne obtains

$$\frac{2}{a} \sum_{k=0}^{n} \mu_k \sum_{L=0}^{[k/2]} \frac{P_k^{(2L)}(T - t)}{a^{2L}} \equiv \sum_{k=0}^{n} \lambda_k P_k(T - t) \tag{41}$$

here

$$P_k^{(2L)}(T - t) \equiv \frac{d^{(2L)}}{dt^{(2L)}} P_k(t) \Big|_{t=T-t}. \tag{42}$$

Multiplying both sides of (41) by $P_i(T - t)$, and integrating over the interval $(0, T)$, one obtains

$$\frac{2}{a} \sum_{k=0}^{n} \mu_k \int_0^T \sum_{L=0}^{[k/2]} \frac{P_k^{(2L)}(T - t)P_i(T - t)}{a^{2L}}\, dt = \frac{\lambda_i T}{2j + 1}. \tag{43}$$

Let

$$V_{k,i} = (2j + 1) \int_0^T \sum_{L=0}^{[k/2]} \frac{P_k^{(2L)}(T - t)P_i(T - t)\, dt}{a^{2L}} \tag{44}$$

$$k, j = 0, 1, 2, \cdots, n,$$

and define the matrix V with elements $V_{k+1,j+1}$. Then

$$\frac{2}{aT} V\mu = \lambda. \tag{45}$$

The mean-square error is given by

$$\theta^2 = \lambda' Q, \tag{46}$$

where the prime indicates the transpose is taken. By substituting (38) and (45) for (46) one may write,

$$\theta^2 = \frac{2}{aT} \{Q' B_{11}' V' Q\}. \tag{47}$$

Eq. (47) may be decomposed into a number of components as follows:

a) the matrix V' may be written

$$V' = TI + T\rho'; \tag{48}$$

b) the matrix B_{11}' may be written by (35) as

$$B_{11}' = S^{-1} + H', \tag{49}$$

where I is an $(n + 1) \times (n + 1)$ identity matrix, and ρ' is defined in Appendix II, so that

$$\theta^2 = \frac{2}{aT} \left\{ \sum_{k=0}^{n} Q_k^2[2k + 1] + Q' \Delta_n Q \right\} \tag{50}$$

where

$$\Delta_n = TS^{-1}\rho' + TH' + TH'\rho'. \tag{51}$$

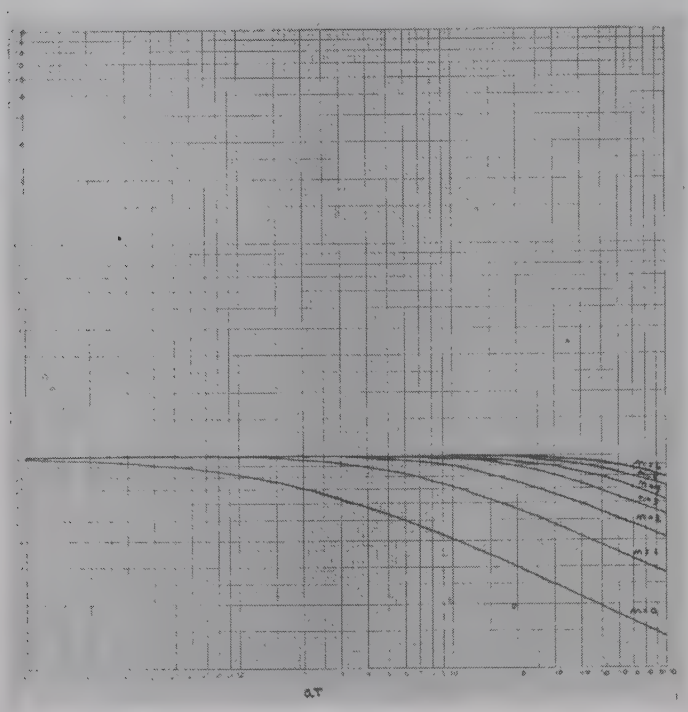
Figs. 1, 2, and 3 present $T^L\theta$ vs aT for $n = 0, 1, 2, \cdots 6$ and $L = 0$, 1, and 2, for $y = 1$, while Figs. 4, 5, and 6 present the same data for $y = 1/2$.

The first term in (50) is the same solution as the white noise case except for the factor $2/a$, and it is independent of aT except for a dependence on T due to Q_k. The second term involves the matrix Δ_n whose elements go to zero as $aT \to \infty$ so that the asymptotic solution is given by

$$\operatorname*{Lim}_{aT\to\infty} [T\theta^2] \cong \frac{2}{a} \sum_{k=0}^{n} Q_k^2[2k + 1]. \tag{52}$$

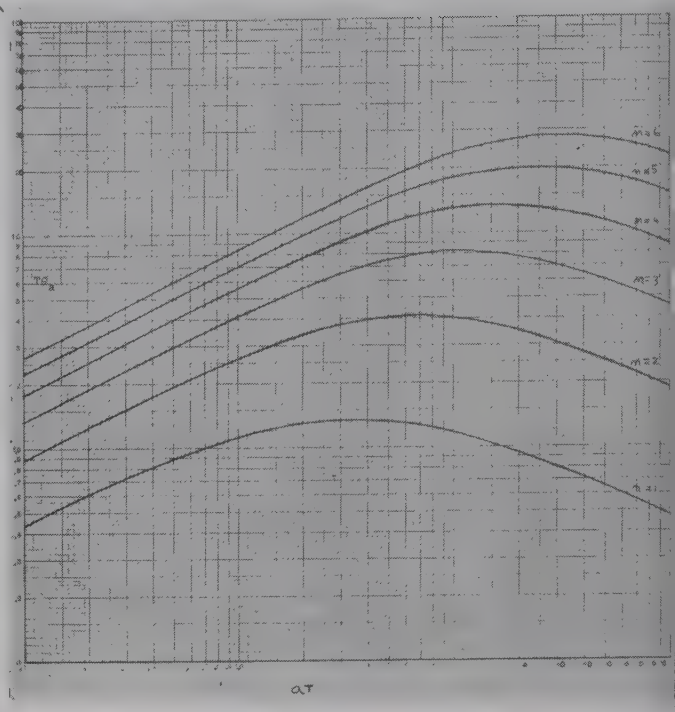
The Q_k defined for the Lth derivative and prediction operator is then

$$[T^{2L+1}\theta^2] \cong \frac{2}{a} H_{n,L}^2(y). \tag{53}$$



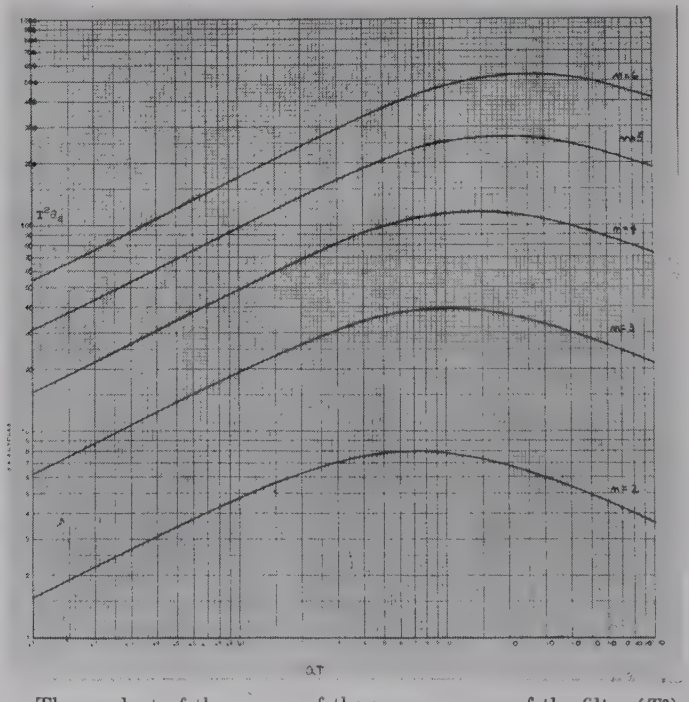
Rms output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$) input noise, parametric in the order (n) of the filter of memory span T. The output is a zero lag ($y = 1$), 0th ($L = 0$) derivative estimate of the input polynomial.

Fig. 1—(θ_a vs aT).



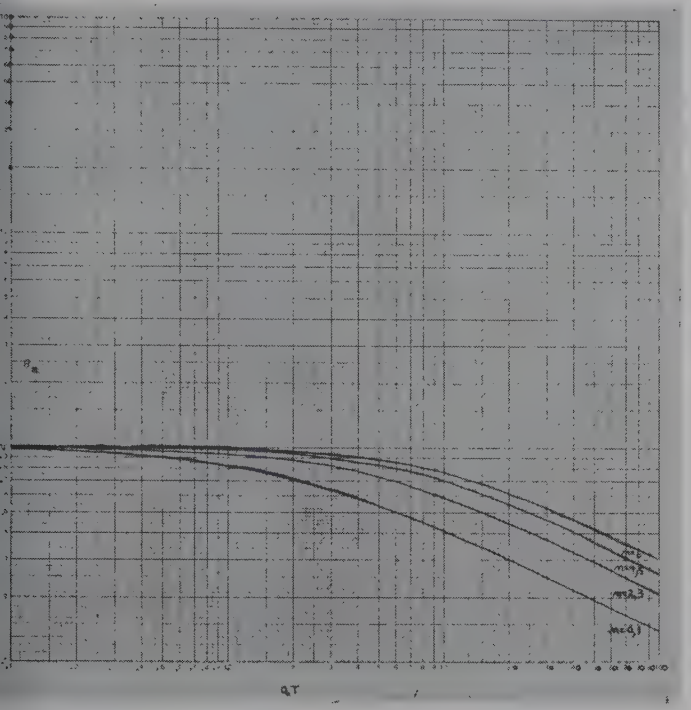
The product of the memory span of the filter (T) and the rm output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$) inpu noise, parametric in the order (n) of the filter. The output is zero lag ($y = 1$), first derivative estimate ($L = 1$) of the inpu polynomial.

Fig. 2—($T\theta_a$ vs aT).



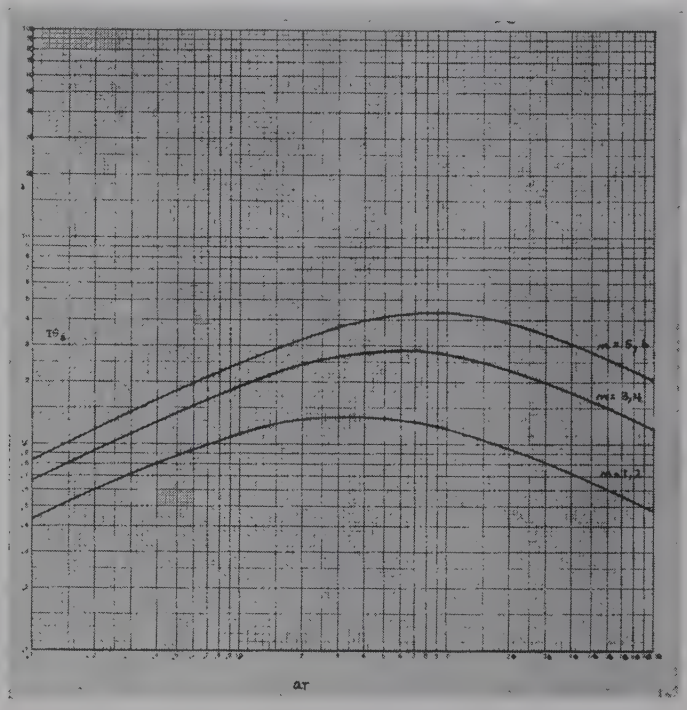
The product of the square of the memory span of the filter (T^2) and the rms output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$) input noise, parametric in the order (n) of the filter. The output is a zero lag ($y = 1$), second derivative ($L = 2$) estimate of the input polynomial.

Fig. 3—($T^2\theta_a$ vs aT).



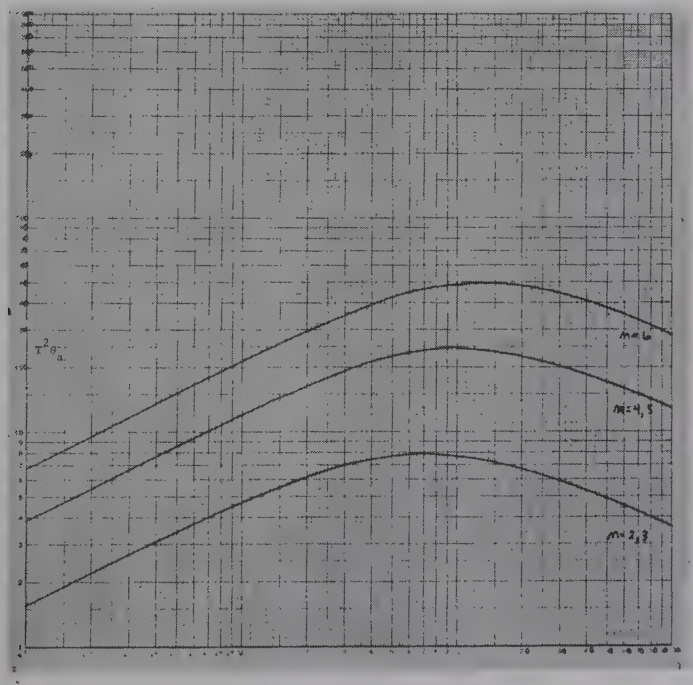
The rms output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$) input noise, parametric in the order (n) of the filter. The output lags the input by 1/2 memory span (midpoint estimate, $y = 1/2$) and is an estimate of the 0th derivative ($L = 1$) of the input polynomial.

Fig. 4—(θ_a vs aT).



The product of the memory span of the filter (T) and the rms output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$), input noise, parametric in the order (n) of the filter. The output lags the input by 1/2 memory span (midpoint estimate, $y = 1/2$) and is an estimate of the first derivative ($L = 1$) of the input polynomial.

Fig. 5—($T\theta_a$ vs aT).



The product of the square of the memory span of the filter (T^2) and the rms output error (θ_a) vs (aT) for exponentially correlated ($e^{-a|\tau|}$), input noise, parametric in the order (n) of the filter. The output lags the input by 1/2 memory span (midpoint estimate, $y = 1/2$) and is an estimate of the second derivative ($L = 2$) of the input polynomial.

Fig. 6—($T^2\theta_a$ vs aT).

Note that the spectra associated with form b) of (19) is given by

$$S(\omega) = \frac{2a}{w^2 + a^2}, \tag{54}$$

so that

$$S(0) = \frac{2}{a}, \tag{55}$$

and (52) may be written for the asymptotic solution

$$\lim_{aT\to\infty}{}^{7}\ \theta^2 \cong S(0)\sigma_\infty^2. \tag{56}$$

The asymptotic solution when form a) of (19) is used, is given by

$$\lim_{aT\to\infty} T\sigma_a^2 \cong \sum_{k=0}^{n} Q_k^2[2k+1], \tag{57}$$

and is identical to the white noise solution (52).

Appendix I

RMS Error Proportionality Factor $H_{n,L}$

A listing of the functions $H_{n,L}^2(z)$, $z = (y - 1/2)^2$ for $n = 0, 1, 2, \cdots 6$, and $L = 0$, 1, and 2 follows. Figs. 7, 8, and 9 contain the graphs of $H_{n,L}(y)$ vs y for the zero, first and second derivative estimators, parametric in the order of the filter (n).

Note that the functions $H_{n,L}(y)$ have maxima and minima in the interval $1/2 \leq y < 1$. For fixed memory span, T, $H_{n,L}(y)$ is directly proportional to the rms output error σ_∞. For even order filters (when the order exceeds the order of the derivative) one obtains the smallest rms errors by estimating the input polynomial at a point other than the mid point, and thus gains by obtaining more accurate estimates with smaller lags. For prediction ($|y| > 1$) note that the rms increases monotonically with the order of the filter (n), as $|y|^{(n-L)}$.

The functions $H_{n,0}^2(z)$, $n = 0, 1, \cdots 6$ are given by

$$H_{0,0}^2(z) = 1,$$

$$H_{1,0}^2(z) = 1 + 12z,$$

$$H_{2,0}^2(z) = 2.25 - 18z + 180z^2,$$

$$H_{3,0}^2(z) = 2.25 + 45z - 660z^2 + 2800z^3,$$

$$H_{4,0}^2(z) = 3.515625 - 56.25z + 1837.5z^2 - 16100z^3 + 44100z^4,$$

$$H_{5,0}^2(z) = 3.515625 + 98.4375z - 3937.5z^2 + 58590z^3 - 343980z^4 + 698544z^5,$$

[7] Preliminary computations show that the asymptotic approximation for the evaluation of the rms holds within a relative error of 5 per cent if $aT \geq 100$. These calculations were performed on end point estimates. Similar computations for mid point estimates yield relative errors of 1 per cent or less.

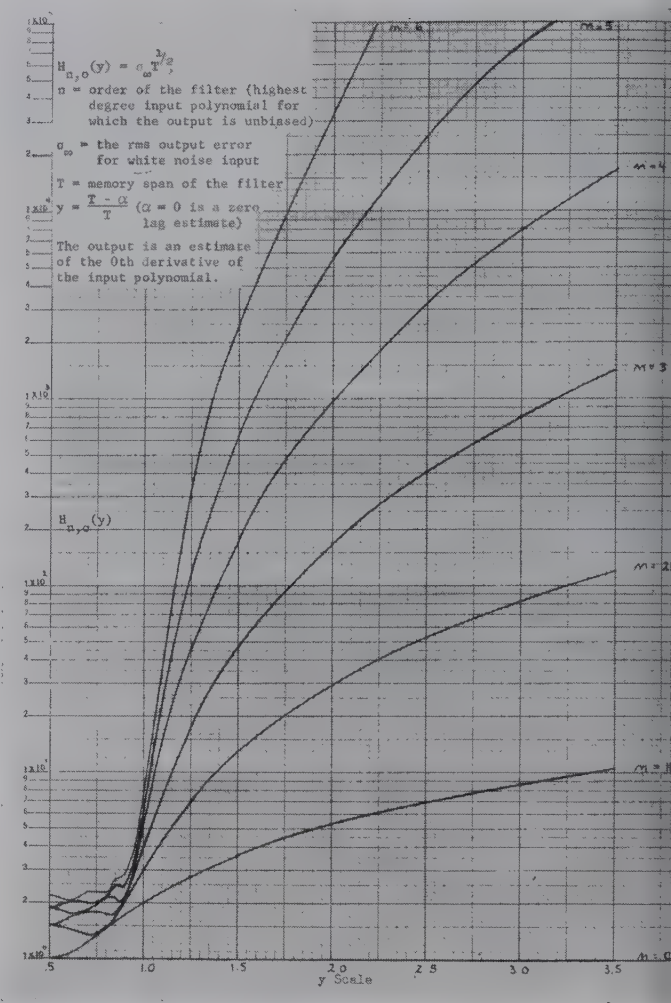


Fig. 7—[$H_{n,0}(y)$ vs y].

and

$$H_{6,0}^2(z) = 4.78515625 - 114.84375z + 7579.6875z^2 - 163905z^3 + 1576575z^4 - 6869016z^5 + 11099088z^6$$

where $z = (y - 1/2)^2$.

The functions $H_{n,1}^2(z)$, $n = 0, 1, 2, \cdots 6$ are given by

$$H_{0,1}^2(z) = 0,$$

$$H_{1,1}^2(z) = 12,$$

$$H_{2,1}^2(z) = 12 + 720z,$$

$$H_{3,1}^2(z) = 75 - 1800z + 25200z^2,$$

$$H_{4,1}^2(z) = 75 + 6300z - 126000z^2 + 705600z^3,$$

$$H_{5,1}^2(z) = 229.6875 - 11075z + 463050z^2 - 5115600z^3 + 17463600z^4$$

and

$$H_{6,1}^2(z) = 229.6875 + 24806.25z - 1256850z^2 + 23090760z^3 - 164157840z^4 + 399567168z^5$$

where $z = (y - 1/2)^2$.

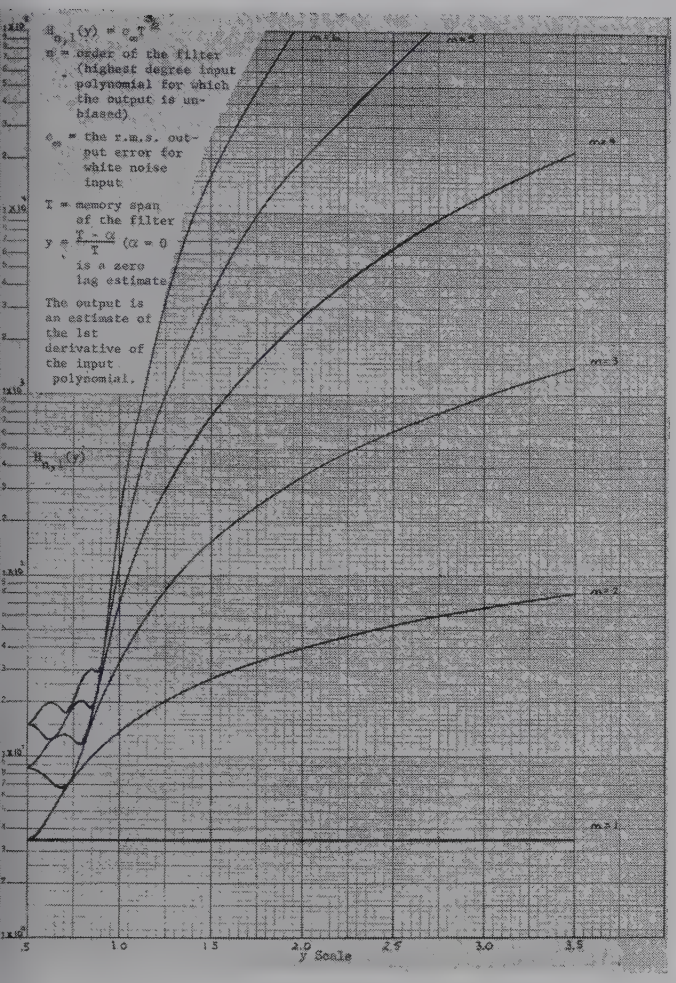


Fig. 8—[$H_{n,1}(y)$ vs y].

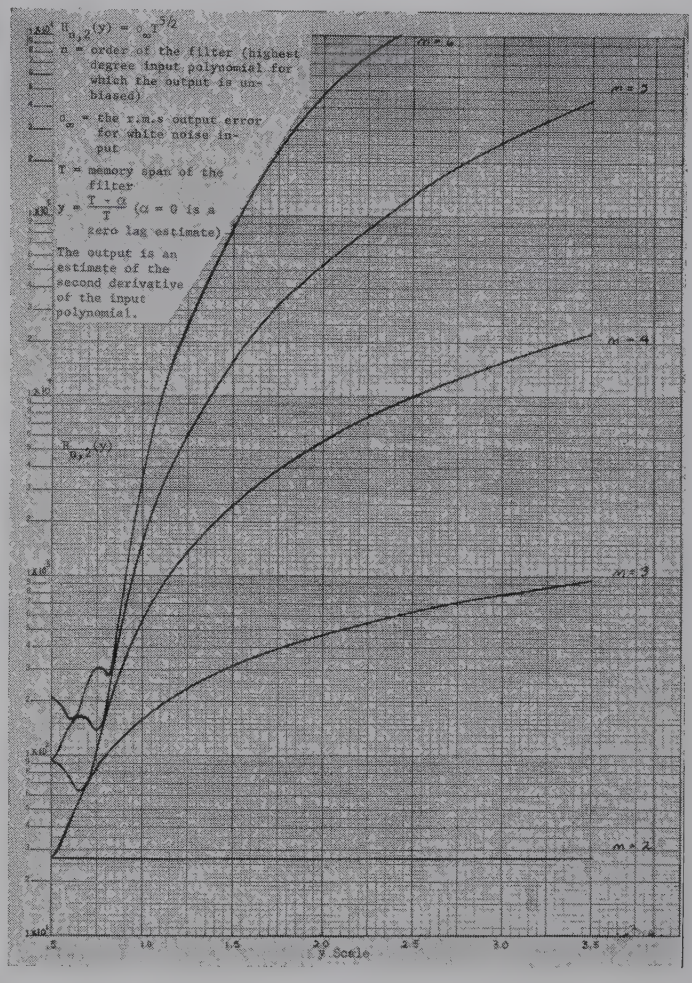


Fig. 9—[$H_{n,2}(y)$ vs y].

The functions $H^2_{n,2}(z)$, $n = 0, 1, 2, \cdots 6$ are given by

$$H^2_{0,2}(z) = 0,$$

$$H^2_{1,2}(z) = 0,$$

$$H^2_{2,2}(z) = 720,$$

$$H^2_{3,2}(z) = 720 + 100800z,$$

$$H^2_{4,2}(z) = 8820 - 352800z + 6350400z^2,$$

$$H^2_{5,2}(z) = 8820 + 1587600z - 40219200z^2 + 279417600z^3,$$

nd

$$H^2_{6,2}(z) = 4461.25 - 3572100z + 183367800z^2 - 2444904000z^3 + 9989179200z^4,$$

here $z = (y - 1/2)^2$.

Appendix II

The matrix V of (45) has elements given by (44) as llows: Let $V_{u,v}$ be the element in the uth row and vth lumn

$$u, v = 1, 2, 3, \cdots, n + 1;$$

then

$$V_{u,v} = 0 \qquad u > v$$

$$V_{u,v} = 0$$

if u and v are not jointly odd or even.

$$V_{u,u} = T,$$

$$V_{1,3} = T \cdot \frac{12}{(aT)^2},$$

$$V_{1,5} = \frac{T}{(aT)^2}\left\{40 + \frac{1680}{(aT)^2}\right\},$$

$$V_{1,7} = \frac{T}{(aT)^2}\left\{84 + \frac{20160}{(aT)^2} + \frac{665280}{(aT)^4}\right\},$$

$$V_{2,4} = \frac{T}{(aT)^2}\{60\},$$

$$V_{2,6} = \frac{T}{(aT)^2}\left\{168 + \frac{15120}{(aT)^2}\right\},$$

$$V_{3,5} = \frac{T}{(aT)^2}\{140\},$$

$$V_{3,7} = \frac{T}{(aT)^2}\left\{360 + \frac{55440}{(aT)^2}\right\},$$

$$V_{4,6} = \frac{T}{(aT)^2}\{252\},$$

and

$$V_{5,7} = \frac{T}{(aT)^2}\{396\}.$$

Let

$$\rho_{u,v} = \frac{V_{u,v}}{T}, \qquad u \neq v,$$

then the matrix $TS^{-1}\rho'$ of (51) is given by, (see 48)

$$TS^{-1}\rho' = \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 5\rho_{1,3} & 0 & 0 & 0 & 0 & \cdots \\ 0 & 7\rho_{2,4} & 0 & 0 & 0 & \cdots \\ 9\rho_{1,5} & 0 & 9\rho_{3,5} & 0 & 0 & \cdots \\ 0 & 11\rho_{2,6} & 0 & 11\rho_{4,6} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{vmatrix}.$$

Note that each element of $TS^{-1}\rho'$ is proportional to $(aT)^{-2}$ for $(aT) \gg 1$. The $(n+1) \times (n+1)$ matrix TH' is given by,

$$TH' = 2\begin{vmatrix} \frac{K_0}{\theta_1^{(n)}} & 0 & \frac{5K_0}{\theta_1^{(n)}} & 0 & \cdots \\ 0 & \frac{3K_1}{\theta_2^{(n)}} & 0 & \frac{7K_1}{\theta_2^{(n)}} & \cdots \\ \frac{K_2}{\theta_1^{(n)}} & 0 & \frac{5K_2}{\theta_1^{(n)}} & 0 & \cdots \\ 0 & \frac{3K_3}{\theta_2^{(n)}} & 0 & \frac{7K_3}{\theta_2^{(n)}} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{vmatrix},$$

where

$$K_j = \frac{B_j}{S_j} \qquad j = 0, 1, 2, \cdots, n$$

$$\theta_i^{(n)} = 1 - 2X_i^{(n)} \qquad i = 1, 2.$$

The functions K_j are given by (5) and (28), and are a proportional to $(aT)^{-1}$ for $aT \gg 1$ so that each elemer of TH' approaches zero as $[aT]$ approaches infinity Finally, the matrix $TH'\rho'$ is obtained and it is seen tha each element of $TH'\rho'$ approaches zero as the $0[aT]^{-1}$ for $[aT] \gg 1$.

A listing of the functions K_j for $j = 0, 1, 2, \cdots 6$ are a follows: Let

$$K_j = \frac{B_j}{S_j},$$

$$K_0 = -\frac{1}{aT},$$

$$K_1 = -\frac{3}{aT}\left\{1 + \frac{2}{aT}\right\},$$

$$K_2 = -\frac{5}{aT}\left\{1 + \frac{6}{aT} + \frac{12}{(aT)^2}\right\},$$

$$K_3 = -\frac{7}{aT}\left\{1 + \frac{12}{aT} + \frac{60}{(aT)^2} + \frac{120}{(aT)^3}\right\},$$

$$K_4 = -\frac{9}{aT}\left\{1 + \frac{20}{aT} + \frac{180}{(aT)^2} + \frac{840}{(aT)^3} + \frac{1680}{(aT)^4}\right\},$$

$$K_5 = -\frac{11}{aT}\left\{1 + \frac{30}{aT} + \frac{420}{(aT)^2} + \frac{3360}{(aT)^3} + \frac{15120}{(aT)^4} + \frac{30240}{(aT)^5}\right\}$$

and

$$K_6 = -\frac{13}{aT}\left\{1 + \frac{42}{aT} + \frac{840}{(aT)^2} + \frac{10080}{(aT)^3} + \frac{75600}{(aT)^4} + \frac{332640}{(aT)^5} + \frac{665280}{(aT)^6}\right\}$$

Some Quantum Effects in Information Channels*

T. E. STERN†, MEMBER, IRE

Summary—**In this paper the quantum nature of electromagnetic adiation is used as a basis for a mathematical model of a continuous hannel. It is shown that this "photon channel" model leads to ore realistic conclusions regarding information transmission. mong the results obtained are:**

1) the maximum entropy for a narrow-band source under an verage power limitation,

2) the frequency distribution (Bose-Einstein) for a wide-band ower-limited source,

3) the transmission rate through a Poisson channel with additive oisson noise.

I. Introduction

IN this paper some of the postulates of information theory, as it applies to continuous channels, will be modified in order to put the theory in closer correspondence with physical laws. It will be shown that these modifications produce new and more useful results in the anges where quantum effects become important.

The definitions of entropy, information rate, channel apacity, etc., as proposed by Shannon[1] for discrete ources and channels, leave little to be desired. They form consistent mathematical system, and, what is more mportant, they are in harmony with our intuitive notions. Iowever, as many investigators (including Shannon) have ointed out, a formalistic extension of the theory to continuous channels leads, at times, to somewhat erroneous onclusions. Consider, for example, the well-known xpression for the capacity of a continuous channel with verage signal power S, average noise power N (white, Gaussian), and bandwidth W,

$$C = W \log \left(1 + \frac{S}{N}\right).$$

s N approaches zero the capacity becomes infinite. ntuitively, such a result must be rejected on the grounds hat it implies a receiver and transmitter with infinite mplitude resolution. However, if we are to reject such a ormulation, what resolution should be assumed? An nswer to this question lies in the dual nature of electromagnetic radiation. Consideration of the wave-like nature f radiation leads naturally to the continuous model for n information channel and to results typified by the bove example. On the other hand, consideration of the orpuscular, or quantized nature of radiation leads aturally to the discrete model. Just as it is necessary to eplace the classical wave model of radiation by the quantum model to explain certain physical phenomena, so too is it possible to utilize the quantum model to give a more realistic picture of information transmission in those cases where quantum effects become important.

Situations in which the quantized nature of the electromagnetic field is dominant are becoming increasingly common. As communications systems move to higher frequencies and spread their power over wider bandwidths, the average number of photons per unit time-bandwidth becomes correspondingly smaller. Thus, it is pertinent for the communication engineer to inquire into the limitations placed upon information rate by the photon nature of radiation. (A discussion of these effects in amplifiers of the Maser type will appear in a forthcoming paper.[2])

Gabor [3,4] was apparently the first to introduce quantum effects into the derivation of information rates, deducing in a somewhat intuitive fashion an expression for channel capacity. His expression resembles Shannon's expression for large signal and noise power. Furthermore, Gabor's work appears to be unique in this area, although many physicists[5,6] have gone in the reverse direction, applying information theoretic principles to problems in modern physics.

In this paper, the quantized nature of the electromagnetic field is used to convert the continuous channel into an equivalent "photon channel," and to derive some expressions for entropy and information rate in such channels. In Section II, the photon channel and source are defined. The maximum entropy distribution for a photon source with average power limitation is derived and discussed in Section III. In Sections IV and V the Poisson channel is considered. It is shown to resemble the continuous Gaussian channel for large average power but to behave quite differently for small average power.

II. The Photon Channel

In constructing a mathematical model for a photon channel, it is necessary to consider two fundamental relations of quantum mechanics: 1) the Planck relation, $E = hf$ (where E is the energy of a photon at frequency f

* Received by the PGIT, November 2, 1959. This work was artially supported by the National Science Foundation Grant SF-G 9780. Publication was assisted by the Marcellus Hartley und.

† Dept. of Elec. Engrg., Columbia University, New York, N. Y.

[1] C. E. Shannon and W. Weaver, "The Mathematical Theory of ommunication," University of Illinois Press, Urbana, p. 67; 1949.

[2] T. E. Stern, "Information rates in photon channels and photon amplifiers," to be published in the 1960 IRE National Convention Record.

[3] D. Gabor, "Lectures on Communication Theory," Mass. Inst. Tech., Res. Lab. of Electronics, Cambridge, Mass., Tech. Rept. No. 238; April 3, 1952.

[4] D. Gabor, "Communication theory and physics," *Phil. Mag.* vol. 41, no. 7, pp. 1161–87; 1950.

[5] L. Brillouin, "Science and Information Theory," Academic Press, Inc., New York, N. Y.; 1956.

[6] D. M. MacKay, "Quantal aspects of scientific information," IRE Trans. on Information Theory, vol. IT-1, pp. 60–80; February, 1953.

and h is Planck's constant), and 2) the uncertainty principle, $\Delta t\ \Delta E \geq h/2\pi$ (here Δt and ΔE represent uncertainty in time and energy respectively). Relation 1) essentially establishes a discrete set of energy levels in the channel, while 2) is automatically satisfied by application of the sampling theorem. This will become apparent as we proceed.

Consider a continuous source transmitting over a bandwidth Δf, with center frequency f_0, such that $\Delta f/f_0 \ll 1$. The sampling theorem states that such a signal has $2\Delta f$ degrees of freedom (DOF) per second. These may be expressed, for example, as amplitude and phase of the carrier at intervals of $1/\Delta f$ seconds. For our purposes, however, it is more instructive to formulate the problem somewhat differently. Consider the channel as being made up of rectangular cells in time-frequency space, of dimensions Δt and Δf, where $\Delta t = 1/2\Delta f$. Each cell represents a DOF of the signal. The source transmits information by locating an arbitrary number of photons in each cell; the receiver is simply a photon counter. It can be seen by using relations 1) and 2) that the uncertainty principle is not violated by such a model. Although there is some question as to how close such a model may be approximated in practice (Gabor, for example, bases his development on the fact that the source cannot determine precisely the number of photons in each cell), the above model serves as a useful point of departure in considering quantum effects. Since the system is assumed narrow-band, all photons have nominally the same energy, $E = hf_0$. Observe that in this model, *amplitude* resolution improves with increasing power level and decreasing frequency.

The photon source can now be defined by a discrete first-order probability distribution over the nonnegative integers. (Only zero memory sources will be considered here.) Having defined the photon channel, we are now in a position to derive expressions for the entropy of photon sources.

III. A Maximum Entropy Photon Source

Consider a source with probability distribution $p(n)$, the probability of locating n photons in a particular DOF. The distribution is assumed identical for each DOF. The entropy for this source is defined as[7]

$$H = -\sum_{n=0}^{\infty} p(n) \ln p(n) \quad \text{(natural units per DOF)}. \tag{1}$$

(For convenience, natural logarithms will be used throughout.)

We now calculate the maximum entropy distribution for a narrow-band source under the average power constraint:

$$\sum_{n=0}^{\infty} np(n) = \bar{N} = \frac{P\ \Delta t}{hf_0} \quad \text{photons/DOF} \tag{2}$$

$$P = \text{average power}.$$

[7] C. E. Shannon, *op. cit.*, see p. 19.

$\bar{N}$, the "occupation number," plays the fundamental ro in the derivations which follow.

Using the additional constraint

$$\sum_{n=0}^{\infty} p(n) = 1, \tag{3}$$

we maximize with respect to $p(n)$ obtaining

$$\frac{\partial}{\partial p(n)}\left[-\sum_{n=0}^{\infty} p(n) \ln p(n) + \lambda \sum_{n=0}^{\infty} np(n) + \mu \sum_{n=0}^{\infty} p(n)\right] = 0 \qquad n = 0, 1, 2, \cdots,$$

where λ and μ are Lagrange multipliers. Solving for $p($ and substituting into (2) and (3), we obtain

$$p(n) = \alpha e^{\lambda n} \quad \text{where} \quad \begin{cases} \lambda = -\ln\left(1 + \dfrac{1}{\bar{N}}\right) \\ \alpha = e^{\mu - 1} = \dfrac{1}{1 + \bar{N}}. \end{cases} \tag{5}$$

Thus, the maximum entropy photon distribution unde the average power constraint is exponential. Substitutin into (1), we obtain for the entropy,

$$\begin{aligned} H &= -\sum_{n=0}^{\infty} \alpha e^{\lambda n} \ln (\alpha e^{\lambda n}) \\ &= -\alpha \sum_{n=0}^{\infty} e^{\lambda n}(\ln \alpha + \lambda n) \\ &= \frac{-\alpha \ln \alpha}{1 - e^{\lambda}} + \frac{\alpha \lambda e^{\lambda}}{(1 - e^{\lambda})^2} \\ &= \bar{N} \ln\left(1 + \frac{1}{\bar{N}}\right) + \ln (1 + \bar{N}). \end{aligned} \tag{6}$$

It is of interest to examine the asymptotic behavior H. For $\bar{N} \gg 1$, we have

$$H \approx \ln e(1 + \bar{N}). \tag{7}$$

As should be expected, this approaches $\ln e\bar{N}$, t maximum entropy of a continuous (exponential) dist bution limited to the positive axis, with mean $\bar{N}$. Shannon has pointed out, the entropy of a continuo distribution has physical meaning only in a relative sen This becomes apparent in comparing the asympto behavior of the discrete and continuous distributions $\bar{N} \to 0$. The discrete expression goes to zero asym totically:

$$H \approx -\bar{N} \ln \bar{N} \to 0 \quad \text{as} \quad \bar{N} \to 0$$

The continuous expression, however, becomes negati for $\bar{N} < 1/e$ and hence is physically meaningless in t range.

Eq. (6) can now be used to derive the photon dist bution in frequency corresponding to a maximum entro wide-band source. As will be shown, this distribution similar to that for thermal noise. (The similarity of t results in this section to statistical mechanical resu

hould be expected since the derivations are completely nalogous to those of statistical mechanics.) Assume that he source is transmitting simultaneously and inde- endently over an infinite number of narrow-band hannels of bandwidth Δf. Let $\bar{N}_i$ = occupation number mean number of photons per DOF) at frequency f_i. Then

$$H_i = \bar{N}_i \ln\left(1 + \frac{1}{\bar{N}_i}\right) + \ln(1 + \bar{N}_i)$$

where H_i = entropy per DOF at frequency f_i.

$$H' = \frac{1}{\Delta t}\sum_{i=1}^{\infty} H_i = 2\,\Delta f \sum_{i=1}^{\infty} H_i \tag{9}$$

where H' = total entropy per unit time.

Expressing the average power constraint as

$$2\,\Delta f \sum_{i=1}^{\infty} \bar{N}_i h f_i = P, \tag{10}$$

we maximize subject to (10) to obtain

$$\frac{\partial}{\partial \bar{N}_i}\left\{\left[2\,\Delta f \sum_{i=1}^{\infty} \bar{N}_i \ln\left(1 + \frac{1}{\bar{N}_i}\right) + \ln(1 + \bar{N}_i)\right] + 2\lambda\,\Delta f \sum_{i=1}^{\infty} \bar{N}_i h f_i\right\} = 0 \qquad i = 1, 2, \cdots \tag{11}$$

nd

$$\ln\left(1 + \frac{1}{\bar{N}_i}\right) + \lambda h f_i = 0$$

$$\bar{N}_i = \frac{1}{e^{h f_i \lambda} - 1}. \tag{12}$$

o evaluate λ we substitute (12) into (10):

$$2\,\Delta f \sum_{i=1}^{\infty} \frac{h f_i}{e^{h f_i \lambda} - 1} = P \tag{13}$$

here $f_i = i\Delta f$. Since Δf may be chosen arbitrarily as long s the relation $\Delta f \Delta t = 1/2$ is satisfied. λ will be evaluated or the limiting case $\Delta f \to 0$.

We have then

$$\lim_{f \to 0} \sum_{i=1}^{\infty} \frac{2h f_i\,\Delta f}{e^{h f_i \lambda} - 1} = \int_0^{\infty} \frac{2hf\,df}{e^{hf\lambda} - 1} = P \qquad \text{(total signal power)}. \tag{14}$$

olving for λ,

$$\lambda = \frac{\pi}{\sqrt{3Ph}}.$$

In order to draw an analogy with thermal radiation, e make the identification

$$\lambda = \frac{1}{kT_e},$$

$$T_e = \frac{1}{k\pi}\sqrt{3Ph},$$

where k = Boltzmann's constant and T_e = effective signal temperature. Substituting into (12),

$$N(f) = \frac{1}{\exp\left(\frac{hf}{kT_e}\right) - 1} \text{ photons/DOF}. \tag{15}$$

Eq. (15) is a Bose-Einstein distribution and is characteristic of thermal radiation.[8] The exponent hf/kT_e in (15) determines whether the system is in the "classical" (continuous) range, or in the quantum range. Specifically,

$\frac{hf}{kT_e} \ll 1$ large occupation numbers (continuous) and

$\frac{hf}{kT_e} \gg 1$ small occupation numbers (quantum).

Since conventional noise theory normally assumes large occupation numbers, no parallel can be drawn to existing formulations for the case of small occupation numbers. However, for the former case, it is of interest to examine the expression for signal power density. From (15) the average number of photons per unit time transmitted in the frequency interval $(f, f + df)$ is

$$N'(f)\,df = 2N(f)\,df = \frac{2\,df}{\exp\left(\frac{hf}{kT_e}\right) - 1} \text{ photons/second}.$$

For the power in this differential bandwidth, we have

$$dP = N'(f) hf\,df = \frac{2hf\,df}{\exp\left(\frac{hf}{kT_e}\right) - 1}, \tag{16}$$

$$\approx 2kT_e\,df \quad \text{for} \quad \frac{hf}{kT_e} \ll 1. \tag{17}$$

Except for the factor of 2, (17) is identical to the Nyquist expression for thermal noise.[9] Going from (16) to (17), the Bose-Einstein distribution was approximated for large occupation numbers by the Boltzmann distribution. It is not surprising that a signal resembling Nyquist ("white") noise should result in this case, since the Nyquist expression is based on the assumption of large occupation numbers. The factor of 2 may be accounted for by noting that in this system we are transmitting all photons in one direction: from the source to the receiver. In a system excited by thermal noise, half the noise power is flowing toward the source and the other half toward the receiver.

An important consequence of the consideration of quantum effects is that (16), the power density spectrum for a maximum entropy source, trails off rapidly to zero at high frequencies. In contrast, a wide-band source based on the conventional model would have a completely uniform spectrum, requiring that there be zero power over any finite frequency range.

[8] W. P. Allis and M. A. Herlin, "Thermodynamics and Statistical Mechanics," McGraw-Hill Book Co., Inc., New York, N. Y., p. 221; 1952.

[9] *Ibid.*, p. 197.

IV. The Poisson Source

In the case of the conventional continuous channel, it is convenient to work with Gaussian distributions for many reasons, two of the most important being: 1) they have maximum entropy under the average power constraint, and 2) they are additive; *i.e.*, a random process which is the sum of two Gaussian processes is also Gaussian. Unfortunately, the exponential photon distribution, which was shown to possess maximum entropy under the average power constraint, is not an additive distribution. The Poisson distribution, however, does have this property. Since additivity is a great convenience in calculating information rates, and since the Poisson distribution has other interesting and useful properties, some characteristics of the Poisson source will be explored in this section.

Consider a photon source with Poisson distribution:

$$p(n) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad n = 0, 1, 2, \cdots \text{(photons/DOF)}. \tag{18}$$

As is well known for this distribution,

$$\text{mean}, \quad \bar{N} = \lambda, \tag{19}$$

$$\text{variance}, \quad \sigma^2 = \lambda \tag{20}$$

and, if $P_3(n)$ is the distribution of the sum of two independent Poisson variables with means λ_1 and λ_2, then

$$P_3 = \frac{\lambda_3^n e^{-\lambda_3}}{n!} \tag{21}$$

where $\lambda_3 = \lambda_1 + \lambda_2$.

The entropy of the Poisson source can be calculated in a straightforward manner as follows:

$$\begin{aligned} H(n) &= -\sum_{n=0}^{\infty} p(n) \ln p(n) \\ &= -e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \left[\ln \frac{1}{n!} + n \ln \lambda - \lambda \right] \\ &= e^{-\lambda} \left[\sum_{n=0}^{\infty} \lambda \left(\frac{\lambda^n}{n!} \right) \right. \\ &\quad \left. - \sum_{n=1}^{\infty} \lambda \ln \lambda \frac{\lambda^{n-1}}{(n-1)!} + \sum_{n=0}^{\infty} \frac{\lambda^n \ln n!}{n!} \right] \\ &= \bar{N} - \bar{N} \ln \bar{N} + e^{-\bar{N}} \left[\frac{\bar{N}^2 \ln 2}{2} \right. \\ &\quad \left. + \frac{\bar{N}^3 \ln 6}{6} + \cdots + \frac{\bar{N}^n \ln n!}{n!} + \cdots \right] \end{aligned} \tag{22}$$

where in the last line, λ has been replaced by $\bar{N}$, to keep the notation for the average number of photons per DOF consistent with Section III.

The entropy for the Poisson source cannot be expressed in closed form. However, most of its useful properties can be deduced by examining its asymptotic behavior for $\bar{N} \ll 1$ and $\bar{N} \gg 1$. We observe from (22) that

$$H(\text{Poisson}) \to -\bar{N} \ln \bar{N} \to H(\text{exponential}) \qquad \text{for } \bar{N} \ll 1. \tag{23}$$

For $\bar{N} \gg 1$ we note the following relation between the Poisson and Gaussian distribution:[10]

$$\sum_{\alpha\lambda^{1/2} < (n-\lambda) < \beta\lambda^{1/2}} \frac{\lambda^n e^{-\lambda}}{n!} \to \Phi(\beta) - \Phi(\alpha) \quad \text{as} \quad \lambda \to \infty \tag{24}$$

where $\Phi(x)$ is the Gaussian cumulative distribution function with unity variance and zero mean.

Eq. (24), a result of the central limit theorem, is essentially a statement of the fact that for sufficiently large λ, the envelope of the Poisson distribution is approximated arbitrarily closely by the Gaussian density function with mean and variance λ. Fig. 1 is a comparison of the two for $\bar{N} = \lambda = 10$. Using Shannon's expression for the entropy of a Gaussian source, we have

$$H(\text{Poisson}) \to 1/2 \ln 2\pi e + 1/2 \ln \bar{N} = H(\text{Gaussian}) \text{ for } \bar{N} \gg 1. \tag{25}$$

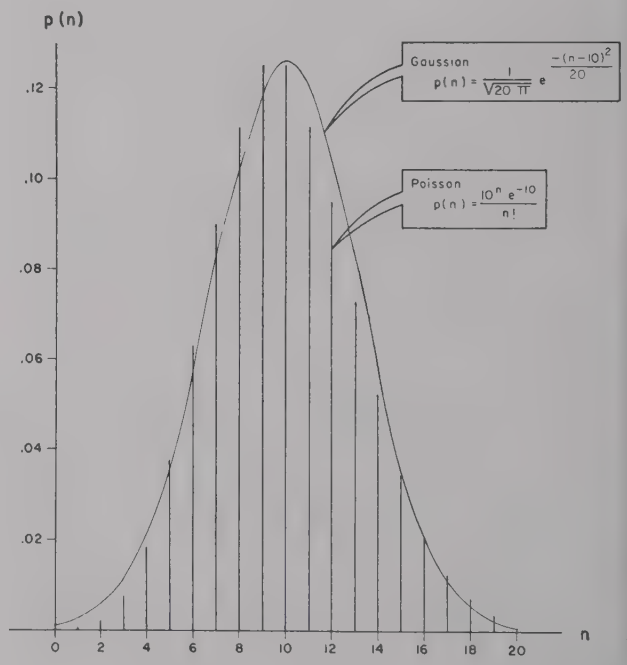


Fig. 1—Comparison of Gaussian and Poisson distributions.

A comparison of the entropies for the exponential, Poisso and Gaussian sources is shown in Fig. 2. (The Gaussia source has a continuous *amplitude* distribution wit mean and variance $\bar{N}$, while the other two are discre photon sources with mean $\bar{N}$.) In Fig. 2, the entropy f the Poisson distribution is approximated by a parti summation of the series to the point where all furth

[10] W. Feller, "An Introduction to Probability Theory and I Applications," John Wiley and Sons, Inc., New York, N. Y., 143; 1950.

erms become negligible according to the computational cheme used. This results in summing approximately $\bar{N} + 7\bar{N})^{1/2}$ terms. Note that the asymptotic expressions iven in (23) and (25) offer excellent approximations for he entropy of the Poisson source in the ranges $\bar{N} \leq 0.4$ nd $\bar{N} \geq 4$. Note also that for large $\bar{N}$ the entropy of he Poisson source becomes approximately one-half the naximum entropy obtainable under the average power onstraint. Since quantum effects have been taken into ccount in the calculation of the entropies of the two iscrete sources, they remain positive for all $\bar{N}$. However, ntropy of the Gaussian source becomes negative for $\bar{N} < 1/2\pi e$.

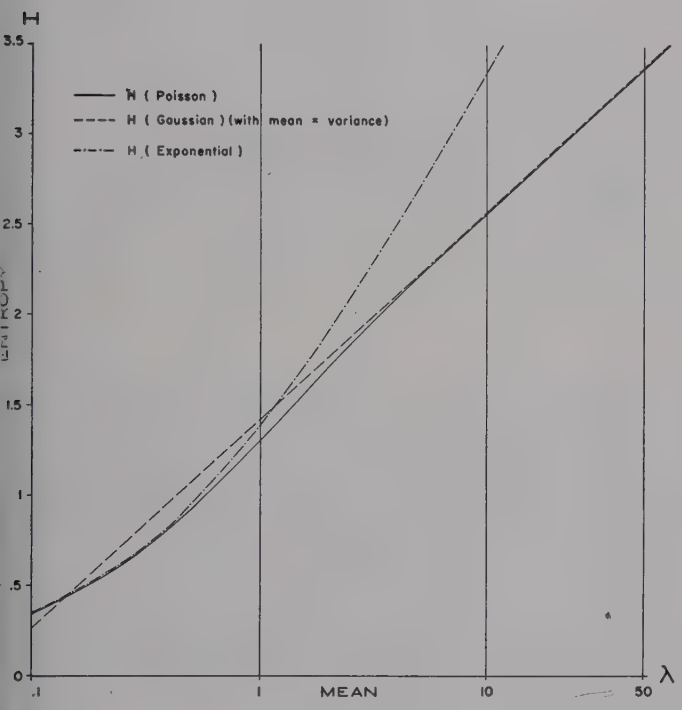


Fig. 2—Entropies as a function of mean, λ.

V. The Poisson Channel

We are now in a position to consider the transmission ate through a photon channel in the presence of additive oise. (A more interesting type of nonadditive noise will e considered in a forthcoming paper.[2]) In order to make ne calculations tractable, it will be assumed here that ne signal and noise are independent Poisson processes ith means, $\bar{N}_s$ and $\bar{N}_n$ respectively. The channel will so be assumed to have a narrow bandwidth Δf. Under ne above conditions, we may express the transmission ate R as

$$= H(\text{Signal} + \text{Noise}) - H(\text{Noise})$$

(Natural units per DOF).

hus

$$= H(\text{Poisson}, \bar{N}_s + \bar{N}_n) - H(\text{Poisson}, \bar{N}_n). \quad (26)$$

Using (25), we may write

$$R \approx 1/2 \ln \left(1 + \frac{\bar{N}_s}{\bar{N}_n}\right) \quad \text{for} \quad \bar{N}_n \gg 1.$$

Since $2\Delta f$ DOF are transmitted per second,

$$R' \approx \Delta f \ln \left(1 + \frac{S}{N}\right) \qquad \text{(natural units per second)} \quad (27a)$$

where S/N, the signal-to-noise *power* ratio has been substituted for $\bar{N}_s/\bar{N}_n$, the signal-to-noise occupation number ratio. Eq. (27a) will be recognized as Shannon's expression for the capacity of the continuous additive Gaussian channel. One obtains quite different results with small occupation numbers, however. A list of some cases of interest follows:

$$R' \approx \frac{\Delta f}{2} \ln [2\pi e(\bar{N}_s + \bar{N}_n)\bar{N}_n^{\bar{N}_n}] \quad \bar{N}_s \gg 1, \bar{N}_n \ll 1, \quad (27b)$$

$$R' \approx \frac{\Delta f}{2} \ln (2\pi e \bar{N}_s) \quad \bar{N}_s \gg 1, \quad \bar{N}_n \to 0, \quad (27c)$$

$$R' \approx \Delta f \ln \left[(\bar{N}_s + \bar{N}_n)^{\bar{N}_s}\left(1 + \frac{\bar{N}_s}{\bar{N}_n}\right)^{\bar{N}_n}\right] \to C$$

$$\bar{N}_s, \bar{N}_n \ll 1 \quad (27d)$$

and

$$R' \to 0 \qquad \bar{N}_s/\bar{N}_n = \text{constant}, \qquad \bar{N}_s \to 0. \quad (27e)$$

It is interesting to review (27a)–(27e) observing how and when they differ from the expression for the analogous continuous channel. A summary of the notable characteristics is given below:

1) R' approaches the continuous expression for large noise occupation numbers. [See (27a).]
2) R' remains finite for all finite signal occupation numbers no matter what the noise power is. [See (27c).]
3) R' goes to zero when the signal occupation number is zero no matter what the noise power is. [See (27e).]
4) R' approaches channel capacity C for small signal and noise occupation numbers. (Assuming the average power constraint.)

We deduce 4) quite easily. Channel capacity C is defined as

$$C = \max_{p(n)} R'$$

where $p(n)$ = source probability distribution.

Since H(Noise) is independent of $p(n)$, channel capacity is attained when H(Signal + Noise) is maximized. Clearly, the maximum is attained when H(Signal + Noise) corresponds to an exponential distribution. But for small occupation numbers, H(Poisson) → H(Exponential); thus the maximum is approached asymptotically in the Poisson channel for small occupation numbers.

VI. Conclusions

In order to circumvent some of the spurious results which are obtained by using the continuum as a physical and mathematical model for an information channel, a discrete photon channel has been postulated. For such a channel, the maximum entropy source under an average power constraint has been shown to have an exponential photon distribution with a Bose-Einstein frequency dependence. By use of the Bose-Einstein statistics, an analogy has been drawn between the maximum entropy source and wide-band thermal radiation at an equivalen temperature T_e. In contrast to continuous sources, th entropy of the discrete source has been shown to be well behaved for small occupation numbers and infinit bandwidths.

The Poisson source has been used to investigate trans mission rate in the presence of additive noise. The result have been shown to approximate those of Shannon for th continuous Gaussian channel in the case of large occu pation numbers, and to be more in keeping with physica limitations for small occupation numbers.

Spectral Analysis of a Process of Randomly Delayed Pulses*

M. V. JOHNS, JR.†

Summary—**Formulas are obtained for the steady-state covariance function of a process of pulses separated by independent random time delays. The pulses considered may be either stochastic or deterministic in character. Since such processes approach stationarity in time their steady-state spectral density functions may be obtained. Explicit results are given for two examples.**

I. Introduction

THE purpose of this paper is to obtain formulas for the covariance and steady-state spectral distribution functions of a process consisting of a sequence of pulses separated by random time delays. Such a process will not be precisely stationary in time but will approach a steady-state condition susceptible to spectral analysis. The wave forms of the pulses are represented by the sample functions of a sequence of independent stochastic processes defined over a finite time interval corresponding to the pulse width, and having identical mean and covariance functions. The time delays are represented by independent identically distributed non-negative random variables possessing a common density function satisfying certain mild restrictions. The derivation of these formulas makes essential use of certain recent developments in renewal theory, a summary of which may be found in [5].

Processes of the type described above arise in communication theory in connection with the analysis of asynchronous multiplexing systems such as those discussed in [4]. In such multiplexing systems, a number of unsynchronized transmitters inject signals into a commo medium to which a similar number of receivers ar coupled. The signal emitted by a typical transmitte consists of a sequence of pulses separated in time b means of a random delay mechanism. Each pulse contain a fragment of the message to be transmitted and possesse features which identify the transmitter. The mathematica model developed in detail in Section II attempts t represent the output of such a transmitter closely enoug for purposes of spectral analysis.

II. The Mathematical Model

The structure of the stochastic process $X(t)$, $t \geq$ whose covariance and asymptotic spectral distributio functions we wish to obtain, is defined in terms of th following quantities:

Let the random variables T_i, $i = 1, 2, \cdots$, be in dependently and identically distributed with commo distribution function $F(t)$, such that $F(0) = 0$ ar $F'(t) = f(t)$ exists, and let $\mu = ET_i < \infty$. The T_i represent the random time delays between successi pulses of the process $X(t)$. Let

$$S_n = \sum_{i=1}^{n} T_i, \qquad n = 1, 2, \cdots; \qquad S_0 = 0, \qquad (1$$

and, for $t \geq 0$, let

$$N(t) = \max \{n: S_n \leq t, S_{n+1} > t\}, \qquad (2$$

$$U(t) = t - S_{N(t)} \qquad (3$$

and

$$V(t) = S_{N(t)+1} - t. \qquad (4$$

* Received by the PGIT, November 10, 1959. Work sponsored in part by the Office of Naval Research, Contract No. N6onr-25140.
† Dept. of Statistics, Stanford University, Stanford, Calif.

The study of the properties of these quantities is a standard part of renewal theory. Let $Y_i(t)$, $0 \leq t \leq c$, $i = 1, 2, \cdots$, be independent stochastic processes representing the successive pulses (of time length c) of the $X(t)$ process. (In particular instances, $Y_i(t)$ might be assumed to be a deterministic function giving the form of the ith pulse, or it might be assumed to be partly stochastic.) We assume that $Y_i(t) = 0$ for t not in the interval $[0, c]$. We further assume that the functions $\xi(t)$ and $\varphi(s, t)$, defined for $s, t \geq 0$ by

$$\xi(t) = EY_i(t) \tag{5}$$

and

$$\varphi(s, t) = EY_i(s)Y_i(t), \tag{6}$$

are finite valued, measurable and independent of i.

We may now represent the $X(t)$ process as follows: For $t \geq 0$,

$$X(t) = Y_{N(t)}[U(t)]. \tag{7}$$

The quantities defined above are illustrated in Fig. 1. This general formulation should be contrasted with that considered by Fortet in [2].

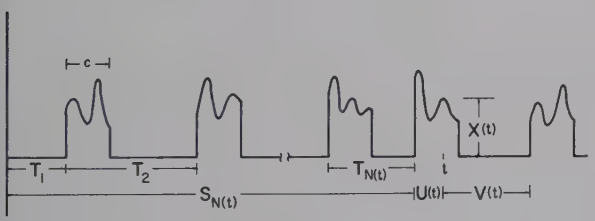


Fig. 1.

II. The Covariance and Spectral Density Functions

We first compute formulas for the covariance function of the $X(t)$ process and then, noting that $X(t)$ is asymptotically stationary, we may compute its asymptotic or steady-state spectral density function. We observe that

$$P\{U(t) \leq u\} = P\{t - S_{N(t)} \leq u\}$$

$$= \begin{cases} \sum_{n=1}^{\infty} P\{t - u \leq S_n \leq t, S_{n+1} > t\}, & 0 \leq u \leq t, \\ P\{S_1 > t\}, & u = t, \end{cases}$$

$$= \begin{cases} \sum_{n=1}^{\infty} \int_{t-u}^{t} [1 - F(t - y)]\, dF^{(n)}(y), & 0 \leq u \leq t, \\ 1 - F(t), & u = t, \end{cases} \tag{8}$$

where $F^{(n)}$ represents the n-fold convolution of F.

Now, for $0 \leq u < t$, let the probability density function of $U(t)$ be given by

$$g_t(u) = \frac{d}{du} P\{U(t) \leq u\}$$

$$= [1 - F(u)] \sum_{n=1}^{\infty} f^{(n)}(t - u), \tag{9}$$

provided that the sum on the right converges uniformly in u (which we will assume), where $f^{(n)}$ represents the derivative of $F^{(n)}$. We note that the probability of the event $U(t) = t$ is the probability that $N(t) = 0$ (*i.e.*, that $T_1 > t$) and is equal to $1 - F(t)$. For $0 \leq u \leq t$, let the conditional probability density function of $V(s)$, given that $U(s) = u$, be given by

$$g(v \mid u) = \frac{d}{dv} P\{V(s) \leq v \mid U(s) = u\}$$

$$= \frac{d}{dv}\left[\frac{F(v + u) - F(u)}{1 - F(u)}\right]$$

$$= \frac{f(v + u)}{1 - F(u)}. \tag{10}$$

Then, letting $g_t(v, u) = g(v \mid u)g_t(u)$, we see that $g_t(v, u)$ represents the joint probability density function for $V(t)$ and $U(t)$ everywhere except for points where $U(t) = t$, for which no density exists. Now

$$E\{X(s)X(s + t)\}$$

$$= E\{X(s)X(s + t) \mid V(s) \leq t\}P\{V(s) \leq t\}$$

$$+ E\{X(s)X(s + t) \mid V(s) > t\}P\{V(s) > t\}. \tag{11}$$

If we designate the first and second terms of the right-hand side of (11) by $E_1(s, t)$, $E_2(s, t)$, respectively, we have

$$E_1(s, t) = E\{E[X(s)X(s + t) \mid U(s), V(s)] \mid V(s) \leq t\}$$

$$\cdot P\{V(s) \leq t\}$$

$$= E\{E[X(s) \mid U(s), V(s)]$$

$$\cdot E[X(s + t) \mid U(s), V(s)] \mid V(s) \leq t\}$$

$$\cdot P\{V(s) \leq t\} \tag{12}$$

since $X(s)$ and $X(s + t)$ are conditionally independent given $U(s)$ and $V(s)$ when $V(s) \leq t$. Furthermore, $X(s)$ and $V(s)$ are conditionally independent given $U(s)$, so that

$$E\{X(s) \mid U(s), V(s)\} = E\{X(s) \mid U(s)\}$$

$$= E\{Y_{N(s)}(U(s)) \mid U(s)\}$$

$$= \xi(U(s)). \tag{13}$$

Similarly, for $V(s) \leq t$,

$$E\{X(s + t) \mid U(s), V(s)\}$$

$$= E\{X(s + t) \mid V(s)\}$$

$$= E\{E[X(s + t) \mid U(s + t), V(s)] \mid V(s)\}$$

$$= E\{E[X(s + t) \mid U(s + t)] \mid V(s)\}$$

$$= E\{E[Y_{N(s+t)}(U(s + t)) \mid U(s + t)] \mid V(s)\}$$

$$= E\{\xi(U(s + t)) \mid V(s)\}. \tag{14}$$

Now, for $0 \leq v \leq t$, $0 \leq x \leq t - v$,

$$P\{U(s + t) \leq x \mid V(s) = v\} = P\{U(t - v) \leq x\}. \tag{15}$$

Hence, from (8), (9) and (14), for $0 \leq v < t$,

$$E\{X(s+t) \mid V(s) = v\}$$

$$= \int_0^{t-v} \xi(x) g_{t-v}(x)\, dx + \xi(t-v)[1 - F(t-v)]. \tag{16}$$

Hence, from (12), (13), (14) and (16),

$$E_1(s,t) = \int_0^s \int_0^t \xi(u) \Big\{ \int_0^{t-v} \xi(x) g_{t-v}(x)\, dx + \xi(t-v)[1 - F(t-v)] \Big\} g_s(v,u)\, dv\, du$$

$$+ \xi(s) \int_0^t \Big\{ \int_0^{t-v} \xi(x) g_{t-v}(x)\, dx + \xi(t-v)[1 - F(t-v)] \Big\} [1 - F(s)] g(v \mid s)\, dv. \tag{17}$$

Now the only case for which $E_2(s,t)$ may be nonzero is when $0 \leq t \leq c$, since otherwise $X(s+t)$ is certainly zero. For $0 \leq t \leq c$, we have

$$\begin{aligned} E_2(s,t) &= E\{E[X(s)X(s+t) \mid U(s), V(s)] \mid V(s) > t\} \\ &\quad \cdot P\{V(s) > t\} \\ &= E\{E[Y_{N(s)}(U(s)) Y_{N(s)}(U(s)+t) \mid U(s)] \\ &\quad \mid V(s) > t\} P\{V(s) > t\} \\ &= E\{\varphi(U(s), U(s)+t) \mid V(s) > t\} \\ &\quad \cdot P\{V(s) > t\} \\ &= \int_0^s \varphi(u, u+t) \int_t^\infty g_s(v,u)\, dv\, du \\ &\quad + \varphi(s, s+t)[1 - F(s)] \int_t^\infty g(v \mid s)\, dv. \end{aligned} \tag{18}$$

From (6) we see that the above expression is automatically zero if $t > c$ so that it holds for all $t \geq 0$. We have now obtained the formulas necessary to compute $EX(s)X(s+t)$ and need only to find an expression for $EX(t)$ in order to be able to compute the covariance function of $X(t)$. But

$$\begin{aligned} EX(t) &= E\{E[X(t) \mid U(t)]\} \\ &= E\{E[Y_{N(t)}(U(t)) \mid U(t)]\} \\ &= E\{\xi(U(t))\} \\ &= \int_0^t \xi(u) g_t(u)\, du + \xi(t)[1 - F(t)]. \end{aligned} \tag{19}$$

It should be noted that a sufficient (although certainly not necessary) condition for the finiteness of the integrals appearing in (17), (18) and (19) is that the functions ξ and φ be bounded.

By definition, the covariance function $R(s,t)$ of $X(t)$ is given by

$$R(s,t) = EX(s)X(t) - EX(s)EX(t). \tag{20}$$

Hence,

$$R(s, s+t) = E_1(s,t) + E_2(s,t) - EX(s)EX(s+t), \tag{21}$$

which may be computed from (17), (18) and (19). For purposes of spectral analysis, however, we wish to take advantage of the asymptotic stationarity of the $X(t)$ process and obtain the limiting covariance function

$$R^*(t) = \lim_{s \to \infty} R(s, s+t). \tag{22}$$

To this end, we first note that a well-known theorem of renewal theory [5] assures us that the so-called "renewal density" $h(x)$ of the T_i's defined by

$$h(x) = \sum_{n=1}^{\infty} f^{(n)}(x) \tag{23}$$

has the property that

$$\lim_{x \to \infty} h(x) = \frac{1}{\mu} \tag{24}$$

provided only that $f(x) \to 0$ as $x \to \infty$ and $|f(x)|^p$ is integrable for some $p > 1$. We assume henceforth that these conditions are satisfied. Now, from (9) we have

$$g_t(u) = [1 - F(u)] h(t-u), \tag{25}$$

so that for each u

$$\lim_{t \to \infty} g_t(u) = \frac{1}{\mu}[1 - F(u)], \tag{26}$$

and, similarly, from (10),

$$\lim_{t \to \infty} g_t(v,u) = \frac{1}{\mu} f(v+u). \tag{27}$$

Hence, noting that the second term of the right-hand side of (17) vanishes for $s > c$, we have

$$\lim_{s \to \infty} E_1(s,t) = \frac{1}{\mu} \int_0^\infty \int_0^t \xi(u) f(v+u) \cdot \int_0^{t-v} \xi(x)[1 - F(x)] h(t-v-x)\, dx\, dv\, du + \frac{1}{\mu} \int_0^\infty \int_0^t \xi(u)\xi(t-v)[1 - F(t-v)] f(v+u)\, dv\, du. \tag{28}$$

Similarly, from (18),

$$\lim_{s \to \infty} E_2(s,t) = \int_0^\infty \varphi(u, u+t) \int_t^\infty \frac{1}{\mu} f(v+u)\, dv\, du = \frac{1}{\mu} \int_0^\infty \varphi(u, u+t)[1 - F(t+u)]\, du, \tag{29}$$

and from (19),

$$\lim_{t \to \infty} EX(t) = \frac{1}{\mu} \int_0^\infty \xi(u)[1 - F(u)]\, du. \tag{30}$$

It is easily verified that a sufficient condition for the validity of the interchange of limits and integration

performed in obtaining the above expressions is again that ξ and φ be bounded functions. Hence, finally, noting that the functions ξ and φ vanish when their arguments fall outside the interval $[0, c]$, we may write

$$R^*(t) = \frac{1}{\mu}\int_0^c \int_0^t \xi(u)f(v+u) \cdot \int_0^{\min(t-v,c)} \xi(x)[1-F(x)]h(t-v-x)\,dx\,dv\,du + \frac{1}{\mu}\int_0^c \int_{\max(0,t-c)}^t \xi(u)\xi(t-v)[1-F(t-v)]f(v+u)\,dv\,du + \frac{1}{\mu}\int_0^{\max(0,c-t)} \varphi(u,u+t)[1-F(t+u)]\,du - \frac{1}{\mu^2}\left\{\int_0^c \xi(u)[1-F(u)]\,du\right\}^2. \tag{31}$$

The asymptotic spectral distribution function $F^{*\prime}$ of the process $X(t)$ is obtained from R^* by means of the usual formula

$$\tfrac{1}{2}[F^*(\lambda+) + F^*(\lambda-)] - F^*(0) = \frac{2}{\pi}\int_0^\infty \frac{\sin 2\pi\lambda t}{t} R^*(t)\,dt, \qquad \lambda > 0, \tag{32}$$

which may be found, for example, in [1]. If it exists, the spectral density function $f^* = F^{*\prime}$ is given by

$$f^*(\lambda) = 4\int_0^\infty R^*(t)\cos 2\pi\lambda t\,dt. \tag{33}$$

It can be shown [3] that the spectrum of an asymptotically stationary process has the same interpretation as in the case of a strictly stationary process.

Example 1: To illustrate the application of the formulas derived in the foregoing section, we consider the particular case where the pulse delay times have exponential distributions and the pulses are rectangular with unit amplitude and time duration c: Let

$$F(t) = \begin{cases} 1 - e^{-t/\mu}, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{34}$$

so that

$$f(t) = \begin{cases} \frac{1}{\mu}e^{-t/\mu}, & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{35}$$

For this case, it is easily verified that

$$h(x) = \frac{1}{\mu}, \qquad x \geq 0. \tag{36}$$

We assume that for each n,

$$Y_n(t) = \begin{cases} 1, & 0 \leq t \leq c \\ 0, & \text{otherwise} \end{cases} \tag{37}$$

with probability one, so that

$$\xi(t) = \begin{cases} 1, & 0 \leq t \leq c \\ 0, & \text{otherwise,} \end{cases} \tag{38}$$

and

$$\varphi(s,t) = \begin{cases} 1, & 0 \leq s, t \leq c \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

Elementary computation shows that for this case (31) becomes

$$R^*(t) = \begin{cases} e^{-c/\mu}(e^{-t/\mu} - e^{-c/\mu}), & 0 \leq t \leq c \\ 0, & t > c, \end{cases} \tag{40}$$

and the corresponding spectral density function given by (33) is

$$f^*(\lambda) = \frac{4e^{-c/\mu}}{1+(2\pi\lambda\mu)^2} \cdot \left\{\mu - \frac{e^{-c/\mu}}{2\pi\lambda}\sin 2\pi\lambda c - \mu e^{-c/\mu}\cos 2\pi\lambda c\right\}. \tag{41}$$

The functions $R^*(t)$ and $f^*(\lambda)$ are shown in Figs. 2 and 3 for the case $c = 1$, $\mu = 5$.

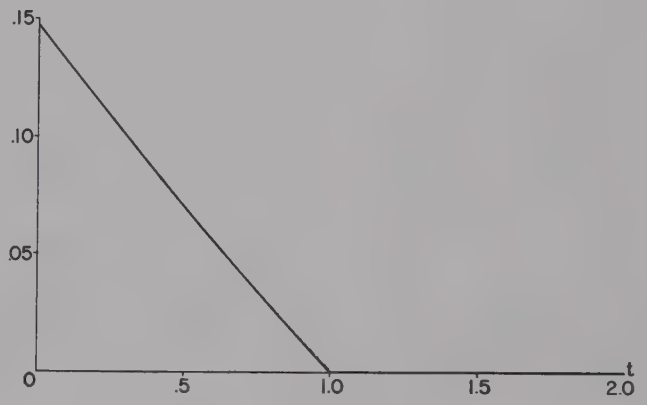


Fig. 2—$R^*(t)$, example 1.

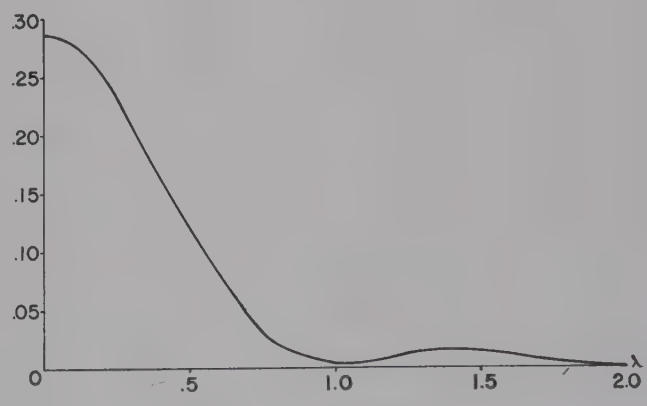


Fig. 3—$f^*(\lambda)$, example 1.

In many cases, the function $h(x)$ cannot be expressed in closed form, so that it is impossible to obtain a closed expression for $R^*(t)$. Another example for which the function $h(x)$ *does* have a closed form expression follows.

Example 2: Suppose that the probability density function for the delay times is given by

$$f(t) = \frac{4}{\mu^2}\, te^{-2t/\mu}. \tag{42}$$

For this case, it is easily verified that

$$h(x) = \frac{1}{\mu}\,(1 - e^{-4x/\mu}). \tag{43}$$

If the quantities $Y_n(t)$, $\xi(t)$ and $\varphi(s, t)$ are as in (37), (38) and (39), then a rather involved computation shows that

$$R^*(t) = \begin{cases} e^{-2(c+t)/\mu}\left(1 + \dfrac{c+t}{\mu}\right) - e^{-4c/\mu}\left(1 + \dfrac{c}{\mu}\right)^2, & t \le c \\ -\dfrac{c^2}{\mu^2}\, e^{-4t/\mu}, & t > c, \end{cases} \tag{44}$$

and

$$\begin{aligned} f^*(\lambda) = {} & \frac{e^{-2c/\mu}}{1 + (\pi\lambda\mu)^2}\left\{\mu + 2c + \frac{2\mu}{1 + (\pi\lambda\mu)^2}\right\} \\ & - \frac{e^{-4c/\mu}\cos 2\pi\lambda c}{1 + (\pi\lambda\mu)^2}\left\{\mu + 4c + \frac{2\mu}{1 + (\pi\lambda\mu)^2}\right\} \\ & - \frac{e^{-4c/\mu}\sin 2\pi\lambda c}{\pi\lambda}\left\{\frac{2c^2}{\mu^2} + \frac{4c}{\mu[1 + (\pi\lambda\mu)^2]} + \frac{2}{[1 + (\pi\lambda\mu)^2]^2}\right\} \\ & + \frac{2c^2 e^{-4c/\mu}}{4 + (\pi\lambda\mu)^2}\left\{\pi\lambda \sin 2\pi\lambda c - \frac{2}{\mu}\cos 2\pi\lambda c\right\}. \end{aligned} \tag{45}$$

The functions $R^*(t)$ and $f^*(\lambda)$ are shown for this example in Figs. 4 and 5 for the case $c = 1$, $\mu = 5$.

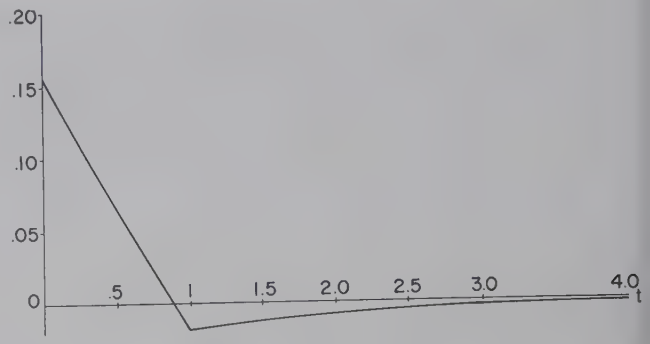


Fig. 4—$R^*(t)$, example 2.

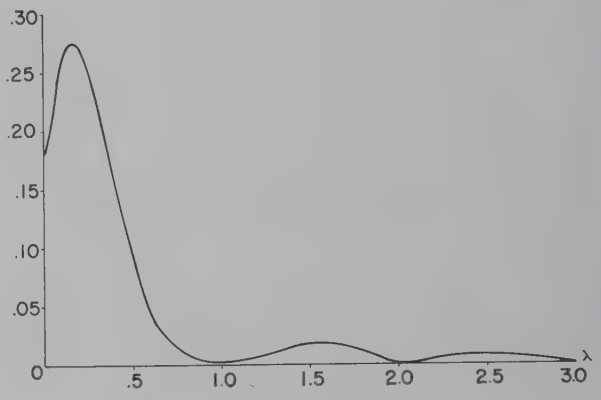


Fig. 5—$f^*(\lambda)$, example 2.

Acknowledgment

The author is indebted to Mrs. Ann Hillier and Mrs. Barbara Miller for performing the calculations necessary for Figs. 2–5.

Bibliography

[1] J. L. Doob, "Stochastic Processes," John Wiley and Sons, Inc New York, N. Y.; 1953.
[2] R. M. Fortet, "Average spectrum of a periodic series of identica pulses randomly displaced and distorted," *Elec. Commun.*, vo 31, no. 4, pp. 283–287; 1954.
[3] E. Parzen, "Asymptotically Stationary Stochastic Processes, (unpublished).
[4] J. R. Pierce and A. L. Hopper, "Nonsynchronous time divisio with holding and with random sampling," Proc. IRE, vol. 4 pp. 1079–1088; September, 1950.
[5] W. L. Smith, "Renewal theory and its ramifications," *J. Roy Statistical Soc.*, ser. B, vol. 20, no. 2, pp. 243–284; 1958.

Binary Codes with Specified Minimum Distance*

MORRIS PLOTKIN†

Summary—Two n-digit sequences, called "points," of binary digits are said to be at distance d if exactly d corresponding digits are unlike in the two sequences. The construction of sets of points, called codes, in which some specified minimum distance is maintained between pairs of points is of interest in the design of self-checking systems for computing with or transmitting binary digits, the minimum distance being the minimum number of digital errors required to produce an undetected error in the system output. Previous work in the field had established general upper bounds for the number of n-digit points in codes of minimum distance d with certain properties. This paper gives new results in the field in the form of theorems which permit systematic construction of codes for given n, d; for some n, d, the codes contain the greatest possible numbers of points.

BY the use of redundancy, it is possible to encode messages for transmission in such a way that errors in transmission may be corrected, provided they are not too dense. For the special case of transmission by means of binary digits, with fixed-length words, this paper investigates the relationships among word length, number of words in the code and number of errors in a word that can be corrected. The best codes known, with respect to these relationships but not to mechanizability, are given in Table I.

In this paper, n-digit binary numbers are regarded as points in an n-dimensional space. The word "point" denotes a binary number, or more accurately, a sequence of binary digits, since the ordinary arithmetical properties of binary numbers are not utilized.[1] Two n-digit points are said to be "at distance d" if they differ in exactly d corresponding digits. For example, the points

1011101000

and

0111001001

are at distance 4, the first, second, fifth, and last digits being different for the two points. A set of n-digit points is called a "code of minimum distance d" if each point is at distance at least d from every other point of the set. The 6-digit points

000000 010101

111000 101101

100110 110011

011110 001011

TABLE I

$A(n, d)$	$n=4$	$n=8$	$n=12$	$n=16$
$d=8$		2 2	2 2 4	4 8 16 32
$d=6$	2	2 2 4	6* 12* 24*	8 16 32
$d=4$	2 2	4 8 16		2^{11}
$d=2$	2 4 8 16 32 64	2^7 2^8	2^9 2^{10} 2^{11}	2^{12} 2^{13} 2^{14} 2^{15}
	2 4 8 16 32 64	2^7 2^8 2^9 2^{10} 2^{11}	2^{12} 2^{13} 2^{14} 2^{15}	2^{16}

* These values differ from the corresponding values of $B(n, d)$.

form a code of minimum distance 3, as may be verified by comparing them pairwise. It is convenient to regard a set consisting of a single point as a code of minimum distance d for every positive integer d.

Clearly, for every ordered pair (n, d) of positive integers there is some maximal number $A(n, d)$[2] of n-digit points which might be selected to give a code of minimum distance d. The code exhibited above demonstrates that $A(6, 3) \geq 8$. It will be seen later than there does not exist a set of nine 6-digit points at a distance 3 or greater pairwise. The $A(n, d)$ notation describes this situation by the equation

$$A(6, 3) = 8.$$

Both the present paper and one by Hamming[1] are concerned primarily with properties of the function $A(n, d)$. [Hamming's paper would not be very different if he had used $A(n, d)$ instead of his $B(n, d)$.] Interest is attached to this function by reason of its connection with coding schemes for correction of errors in systems employing binary symbols for handling information. Consider a system for computing or transmitting n-digit binary numbers, and having the property that noise or system malfunction will affect at most x of the n-digits in any output number. There can be selected $A(n, 2x + 1)$ but no more n-digit numbers which form a code of minimum distance $2x + 1$. If the system can be designed or its operation programmed in such a manner that correct—*i.e.*, error free—operation will give rise to outputs consisting exclusively of numbers in the code, then the correct outputs will always be deducible from the actual output. There will always be exactly one code number at distance x or less from an output number. For example,

* Received by the PGIT, November 20, 1959. The work leading to this paper was sponsored by the Burroughs Corp.

† Auerbach Electronics Corp., Philadelphia, Pa.

[1] R. W. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950.

[2] This definition for $A(n, d)$ is not the same as Hamming's definition for his $B(n, d)$, in that a somewhat less restrictive class of codes is used here. $B(n, d) \leq A(n, d)$ for all n, d. $B(n, d)$ is always a power of 2; $A(n, d)$ need not be. The departure is for convenience only and does not constitute a significant innovation.

if $x = 1$ and $n = 6$, there could be used the code exhibited above to demonstrate that $A(6, 3) \geq 8$, since $d = 3 = 2x + 1$. An output of, for example, 101001 in such a system could be "corrected" to 101101, that being the code number at distance 1 or less from the actual output.

Following is a summary of Hamming's results, which are utilized in the present paper:[3]

$$A(n, 1) = 2^n,$$

$$A(n, 2) = 2^{n-1},$$

$$A(n + 1, 2k) = A(n, 2k - 1),$$

$$A(n, 2k - 1) \leq \frac{2^n}{1 + C(n, 1) + \cdots + C(n, k - 1)}$$

where

$$C(n, h) = \frac{n!}{h!\,(n - h)!}.$$

Except for the unimportant difference between $A(n, d)$ and $B(n, d)$, all definitions and results to this point are due to Hamming.

Definition: By the sum $a * b$ of two n-digit points a, b is meant that n-digit point whose jth digit is zero / unity according as the jth digits of a, b are the same / different.

For example,

if a is 1011101000

and b is 0111001001

then $a * b$ is 1100100001.

For any a, $a * a$ is the origin or null-point $00 \cdots 0$, denoted throughout by o.

Definition: $|a| = m$ means that exactly m of the digits of the point a are 1. In this notation, the distance between two n-digit points a, b is $|a * b|$.

It is clear that addition as defined above is associative: that $(a * b) * c = a * (b * c)$. If K is a code of n-digit points $a, b, c, \cdots$ of minimum distance d, then so is the code denoted by $K * x$ consisting of the points $a * x$, $b * x$, $c * x, \cdots$ where x is any n-digit point. For pairwise distances are preserved, since

$$|\,(a * x) * (b * x)\,| = |\,(a * b) * (x * x)\,|$$

$$= |\,(a * b) * 0\,| = |\,a * b\,|.$$

Theorem 1: If $2d > n$, then $A(n, d) \leq 2m \leq 2d/(2d - n)$, m an integer.

Proof: Let K be any code consisting of A n-digit points of minimum distance d. Let h be any point in K. Consider the code $K * h$, as defined by the notation of the preceding paragraph. Since $h * h = o$, o will be a member of $K * h$. By the minimum distance property it follows that the other $A - 1$ members of $K * h$ each have at least d digits equal to 1, so that the sum of $|k * h|$ over all k in K is at leas $(A - 1)d$. This is true for each of the A possible choice of h. Letting h also run through all possible values we find that the total number N of 1's in the A^2 possible sums o two points $k * h$ for h, k both in K, must satisfy

$$A(A - 1)d \leq N = \sum |h * k|,$$

the sum over all ordered pairs (h, k).

Next, we obtain another inequality on N by considering corresponding digits of each point in K. Suppose x points in K have for their first digit 1 and the other $A - x$ have fo their first digit 0. In the A^2 sums $k * h$, exactly $2x(A - x)$ will have for their first digit 1. If $y, z, \cdots$ are defined i similar manner for the second digits, third digits, $\cdots$ o the points in K, the same number $N = \sum |h * k|$ is seen to be expressible as

$$N = 2x(A - x) + 2y(A - y) + 2z(A - z) + \cdots .$$

Case 1): $A = 2m$. Each of the terms

$$2x(A - x),\ 2y(A - y),\ \text{etc.},$$

is at most $A^2/2$. Since there are n such terms,

$$N \leq nA^2/2.$$

Combining this inequality with $A(A - 1)d \leq N$,

$$2(A - 1)\,d \leq An$$

and

$$(2\,d - n)A \leq 2\,d.$$

Since $2d > n$ by hypothesis,

$$A = 2m \leq \frac{2\,d}{2\,d - n}.$$

Case 2): $A = 2m - 1$. Each of the terms

$$2x(A - x),\ 2y(A - y),\ \text{etc.},$$

is at most $(A^2 - 1)/2$. Continuing as in Case 1), it ma be seen that

$$A = 2m - 1 < 2m \leq \frac{2\,d}{2\,d - n}.$$

Corollary: $A(n, n) = 2$. By the above theorem $A(n, n) \leq 2$, and the pair $00\cdots0$, $11\cdots1$ constitute a example showing $A(n, n) \geq 2$.

Corollary: $A(4m - 1, 2m) \leq 4m$ and $A(4m - 2, 2m \leq 2m$.

Theorem 2: $A(n, d) \leq 2A(n - 1, d)$.

Proof: Let K be a code of $A(n, d)$ n-digit points minimum distance d. Separate the points of K into tw sets according to their first digit. At least one of th two sets will contain one-half or more of the points. D letion of the first digit in each point of that set leaves code of minimum distance d containing at least

$$\frac{A(n, d)}{2}$$

$(n - 1)$-digit points. This proves the theorem.

[3] Hamming's proofs that the relations hold for $B(n, d)$ are valid without change for $A(n, d)$.

Corollary: Since $A(4m - 1, 2m) \leq 4m$, we have

$$A(4m, 2m) \leq 8m.$$

Also, if $A(4m, 2m) = 8m$, then $A(4m - 1, 2m) = 4m$ and

$$A(4m - 2, 2m) = 2m.$$

Theorem 3: If $4m - 1$ is a prime, then $A(4m, 2m) = 8m$.

Proof: Since we have shown $A(4m, 2m) \leq 8m$, it is sufficient to construct a code of $8m$ $4m$≡digit points of minimum distance $2m$.[4] One such construction is included in the Appendix.

Theorem 4: $A(2n, 2d) \geq A(n, 2d)\, A(n, d)$.

Proof: For this proof only the symbol $\frown$ denoting concatenation of two sets of symbols is introduced. Its meaning is illustrated by the example:

If a is 1011

and b is 00111,

then $\widehat{ab}$ is 101100111.

Clearly, $|\widehat{ab}| = |a| + |b|$ for any a, b.

Let L be a code of minimum distance d containing $A(n, d)$ n-digit points, and M a code of minimum distance $2d$ containing $A(n, 2d)$ n-digit points. From these will be constructed a code K of minimum distance $2d$, containing $A(n, d)A(n, 2d)$ points. This will prove the theorem.

Define K as the set of all points $u = \widehat{(aa)} * \widehat{(bo)}$ for a in L and b in M, o being the n-digit null point. The points u will of course be $2n$-digit points. There are $A(n, d)$ $A(n, 2d)$ distinct pairs a, b. If it can be shown that two distinct pairs a_1, b_1 and a_2, b_2 give rise to points u_1, u_2, in K at a distance at least $2d$, the theorem is proved.

$$\begin{aligned} u_1 * u_2 &= \widehat{(a_1\, a_1} * \widehat{b_1\, o)} * \widehat{(a_2\, a_2} * \widehat{b_2\, o)} \\ &= \widehat{(a_1\, a_1} * \widehat{a_2\, a_2)} * \widehat{(b_1\, o} * \widehat{b_2\, o)} \\ &= \{\widehat{(a_1 * a_2)}\, \widehat{(a_1 * a_2)}\} * \{\widehat{(b_1 * b_2)}\, o\}. \end{aligned}$$

For a_1, b_1 and a_2, b_2 distinct pairs, three cases may occur.

1) $a_1 = a_2$, $b_1 \neq b_2$. In this case

$$\begin{aligned} |\, u_1 * u_2 \,| &= |\, \widehat{o\, o} * \{\widehat{(b_1 * b_2)}\, o\} \,| \\ &= |\, \widehat{(b_1 * b_2)}\, o \,| = |\, b_1 * b_2 \,|. \end{aligned}$$

But, by hypothesis, b_1, b_2 are members of code M of minimum distance $2d$, so that in this case $|u_1 * u_2| \geq 2d$.

2) $a_1 \neq a_2$, $b_1 = b_2$. In this case

$$\begin{aligned} |\, u_1 * u_2 \,| &= |\, \{\widehat{(a_1 * a_2)}\, \widehat{(a_1 * a_2)}\} * \{\widehat{o\, o}\} \,| \\ &= |\, \widehat{(a_1 * a_2)}\, \widehat{(a_1 * a_2)} \,| \\ &= 2\, |\, a_1 * a_2 \,| \end{aligned}$$

and since $|a_1 * a_2| \geq d$ by hypothesis, again $|u_1 * u_2| \geq 2d$.

3) $a_1 \neq a_2$, $b_1 \neq b_2$. In this case we write

$$\begin{aligned} |\, u_1 * u_2 \,| &= |\, \widehat{\{(a_1 * a_2) * (b_1 * b_2)\}}\, \{(a_1 * a_2) * o\} \,| \\ &= |\, (a_1 * a_2) * (b_1 * b_2) \,| + |\, a_1 * a_2 \,|. \end{aligned}$$

For any x, y, $|x * y| \geq |x| - |y|$, or

$$|\, x * y \,| + |\, y \,| \geq x.$$

Therefore, $|u_1 * u_2| \geq |b_1 * b_2| \geq 2d$.

Theorem 5: If $A(4m, 2m) = 8m$ holds for $m = x$, then it also holds for $m = 2x$.

Proof: $A(8m, 4m) \geq A(4m, 4m)A(4m, 2m) = 2A(4m, 2m)$. Also, $A(8m, 4m) \leq 16m$ by the corollary to Theorem 2. Therefore, $A(4m, 2m) = 8m$ implies $A(8m, 4m) = 16m$, which was to be proved.

Theorems 3 and 5 together prove that $A(4m, 2m) = 8m$ holds for a number of values of m. In $m \leq 20$, the values which are not reached are 7, 9, 13, 14, and 19. I know of no m for which I can show $A(4m, 2m) \neq 8m$. This suggests the conjecture $A(4m, 2m)$, $= 8m$ for all m.

From Theorems 1 to 5, in conjunction with Hamming's results, there may be deduced for a number of n, d the exact value of $A(n, d)$. With few exceptions, these n, d lie in the region $2d > n$. This is the region in which $A(n, d) \leq 4d$ and one would expect it for that reason to be the least interesting region from the point of view of practical applications. The known values of $A(n, d)$ for $d \leq 8$, $n \leq 16$ are shown in Table I. Corresponding values of $B(n, d)$ are given in Table II. The bottom two rows, $d = 1$ and $d = 2$, are given by Hamming; and values for n, $d = 3, 3$; $4, 4$; $7, 3$; $8, 4$; $15, 3$; $16, 4$ are special cases of Golay's formula.[5] No single method can be

TABLE II

$B(n, d)$															
$d = 8$	.	.	.	.	.	.	.	.	2	2	2	2	4	4	8	16	32
$d = 6$	.	.	.	.	.	.	2	2	2	2	4	4	8	16			
$d = 4$	.	.	.	2	2	2	4	4	8	16							
	.	.	2	2	4	8	16	16	32	64	2^7	2^8	2^9	2^{10}	2^{11}		
$d = 2$	.	2	4	8	16	32	64	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	
	2	4	8	16	32	64	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	
				.			.			.			.			.	
				.			.			.			.			.	
				$4 = n$			$8 = n$			$12 = n$						$16 = n$	

[4] Since the original writing of this report, the author has learned that such codes are a special case of a more general class that may be constructed by methods given by R. E. A. C. Paley, "On Orthogonal Matrices," *J. Math. and Phys.*, vol. 12, pp. 311–320; 1933. By virtue of Paley's work, Theorem 3 may be stated not only for $4m - 1$ prime, but for $4m - 1$ of the form $2^k(p^h + 1)$, p an odd prime and h, k integers.

[5] M. J. E. Golay, "Notes on digital coding," PROC. IRE, vol. 37, p. 657; June, 1949.

prescribed for finding the values given for different n, d. To illustrate the procedures it will be shown that $A(13, 8) = 4$.

Because $8 - 1 = 7$ is a prime, $A(8, 4) = 16$ by $A(4m, 2m) = 8m$. This in turn implies that $A(8m, 4m) = 16m$, or $A(16, 8) = 32$.

$$A(13, 8) \geq \frac{A(14, 8)}{2} \geq \frac{A(15, 8)}{4} \geq \frac{A(16, 8)}{8} = 4$$

by Theorem 2. By Theorem 1,

$$A(13, 8) \leq 2m \leq \frac{16}{16 - 13}, \quad \text{or} \quad A(13, 8) \leqq 4.$$

Combining the two inequalities, $A(13, 8) = 4$.

For $n > 2d$, although they do not provide exact values of $A(n, d)$, Theorems 1 to 5 may be useful in obtaining bounds. Again, the method chosen will be different for different n, d. For purposes of illustration the case $n = 26$, $d = 6$ is discussed.

$$A(26, 6) = A(25, 5) \leq \frac{2^{25}}{1 + 25 + (1/2)(25)(24)} = \frac{128}{163} \cdot 2^{17},$$

$$A(26, 6) \geqq A(13, 6)A(13, 3) \geqq A(12, 6)A(14, 4)$$

$$\geq 24A(7, 4)A(7, 2),$$

and

$$A(26, 6) \geq (24)(8)(64) = 3 \cdot 2^{12}.$$

This tells us that

$$3 \cdot 2^{12} \leq A(26, 6) \leq \frac{128}{163} \cdot 2^{17}.$$

Further, it tells us how to construct a code of 3.2^{12} points for $n = 26$, $d = 6$, because all theorems of this paper bounding $A(n, d)$ from below are constructive in nature. In order to construct such a code the inequalities leading to $A(26, 6) \geq 3.2^{12}$ are retraced.

First let K_1 be the code of 8 points for $n = 4$, $d = 2$, consisting of all 4-digit points which have an even number of 1's among their digits. From K_1 and the two point code 1111,0000, there may be constructed by the method of Theorem 4 a code K_2 of $(8)(2) = 16$ points with $n = 8$, $d = 4$. If the sixteen points of K_2 are separated into two sets according to whether the last digit is 0 or 1, at least one of the sets will have eight or more points and deletion of the last digit will give a code K_3 of at least eight members with $n = 7$, $d = 4$. We have now got as far as $A(7, 4) = 8$ in retracing the inequalities. Next we take K_4 as the code of 64 points consisting of all 7-digit points with an even number of 1's among their digits. K_4 exemplifies $A(7, 2) = 64$. From K_3 and K_4 there may be constructed by the method of Theorem 4 a code K_5 of $A(7, 2)A(7, 4) = 2^9$ points, with $n = 14$ and $d = 4$. By merely suppressing the last digit of K_5 we get a code K_6 with the same number 2^9 of points, $n = 13$, $d = 3$.

Since $4 \cdot 3 - 1 = 11$ is a prime, the method of Theorem 3 permits construction of a code K_7 exemplifying $A(12, 6) = 24$. By the possibly wasteful process of adding an 0 at the end of each point of K_7 there may be constructed K_8, a code of 24 points with $n = 13$, $d = 6$. Finally from K_6 and K_8 there may be obtained, again by the method of Theorem 4, our desired code for $n = 26$, $d = 6$, with at least $24 \cdot 2^9 = 3 \cdot 2^{12}$ points.

Appendix

Proof That $A(4m, 2m) = 8m$ if $4m - 1$ is a Prime

In this proof of congruences are modulo $4m - 1$ if not otherwise noted. The first and greater part of the proof consists of constructing a set $a_1, a_2, \cdots, a_{4m-1}$ of $(4m - 1)$-digit points satisfying

$$| a_i | = 2m \quad \text{and} \quad | a_i * a_k | = 2m, \qquad k \neq j$$

and

$$j, k = 1, 2, \cdots, 4m - 1.$$

After that the rest is simple.

It is a well-known theorem in elementary number theory that every odd prime p has a primitive root r: an integer such that each of $r, r^2, r^3, \cdots, r^{p-1}$ is congruent to a different one of $1, 2, 3, \cdots, p - 1$. Let r be a primitive root of the prime $4m - 1$.

An integer $x \not\equiv 0$ modulo p is called a quadratic residue of a prime p if there exists another integer y satisfying $y^2 \equiv x$ modulo p. If there exists no y satisfying $y^2 \equiv x$ modulo p then x is called a quadratic nonresidue of p. It is known from elementary number theory that exactly half of the integers $1, 2, \cdots, p - 1$ are quadratic residues and half are quadratic nonresidues and that -1 is a quadratic nonresidue of all prines of the form $4m - 1$.

The numbers $r^2, r^4, \cdots, r^{4m-2}$ are all quadratic residues of $4m - 1$, for clearly $y = r^k$ satisfies $y^2 \equiv r^{2k}$ for $k = 1, 2, \cdots, 2m - 1$. Therefore, $r, r^3, \cdots, r^{4m-3}$ must be quadratic nonresidues of $4m - 1$. The numbers $-r^2, -r^4, \cdots, -r^{4m-2}$ are also quadratic nonresidues, for if there were a w satisfying $w^2 \equiv -r^{2k}$ there would be a y—namely, the y satisfying $w = yr^k$—which satisfies $y^2 \equiv -1$, and this is known to be impossible modulo $4m - 1$. The numbers $r, r^3, \cdots, r^{4m-3}$ are, therefore, each congruent to a different one of $-r^2, -r^4, \cdots, -r^{4m-2}$, each set containing one member congruent to each of the nonresidues among $1, 2, 3, \cdots, 4m - 2$.

I shall construct the $a_1, a_2, \cdots, a_{4m-1}$ in terms of their binary digits. To that end, I first define a binary digit z_i for all integral i by:

$z_i = 1$ if i is a quadratic residue of $4m - 1$; *i.e.*, if i is congruent to one of $r^2, r^4, r^6, \cdots, r^{4m-2}$.

$z_i = 0$ if i is a quadratic nonresidue of $4m - 1$; *i.e.*, if i is congruent to one of $r, r^3, \cdots, r^{4m-3}$.

$z_i = 1$ if $i \equiv 0$.

The z_i so defined have the property, as is easily verified, that $z_i = z_{ir^2}$ for every i. It follows that $z_i = z_{ir^2} = z_{ir^4} = z_{ir^6} = \cdots$ etc., for every i. Also, since $r^2, r^4, r^6, \cdots, r^{4m-2}$ are congruent to $-r, -r^3, \cdots, -r^{4m-3}$ in some order the above equations may be expressed

$$z_i = z_{-ir} = z_{-ir^3} = z_{-ir^5} = \cdots \text{ etc., or}$$

$$z_{-i} = z_{ir} = z_{ir^3} = z_{ir^5} = \cdots \text{ etc., for every } i.$$

These equations may all be combined into

$$z_{ir^k} = \begin{cases} z_i, & k \text{ even} \\ z_{-i}, & k \text{ odd} \end{cases}$$

for any i and k.

The a_i are now defined. Let a_1 be the $(4m - 1)$-digit point whose ith digit is z_i, $i = 1, 2, \cdots, 4m - 1$. For $j = 2, 3, \cdots, 4m - 1$, a_j is obtained by cyclic permutation of the digits of a_1: let z_{i+j-1} be the ith digit of a_j. Consider the digits of a_1. The last one is z_{4m-1}, which is 1 because $4m - 1 \equiv 0$. Of the others, half of the subscripts are residues and half are nonresidues; *i.e.*, half the digits are 1's and half 0's. Therefore, $2m$ of the digits are 1's, and $2m - 1$ 0's; $|a_1| = 2m$. But since a_j is obtained by permuting the digits of a_1, we have

$$|a_j| = 2m, \quad j = 1, 2, \cdots, 4m - 1.$$

This is one of the two conditions we shall require on the a_j, the other being

$$|a_j * a_k| = 2m \quad \text{if} \quad j \neq k.$$

We now verify that the second condition is also met.

Let

$$s_{j,k} = |a_j * a_k|, \qquad i, j = 1, 2, \cdots, 4m - 1.$$

I wish to show $s_{jk} = 2m$ for all $j \neq k$. By the cyclic construction of the a_j, it is clear that $|a_j * a_k| = |a_1 * a_{k-j+1}|$ if $k > j$. It is, therefore, sufficient to prove $s_{1,k} = 2m$, $k = 2, 3, \cdots, 4m - 1$; or $s_{1,u+1} = 2m$, $u = 1, 2, \cdots, 4m - 2$.

$s_{1,u+1} = |a_1 * a_{u+1}|$ is investigated directly by comparison of corresponding digits of a_1 and a_{u+1}. These are, in order, the pairs $z_1, z_{u+1}; z_2, z_{u+2}; \cdots; z_{4m-1}, z_u$. $s_{1,u+1}$ is the number of pairs z_j, z_{u+j} for which $z_j \neq z_{u+j}$. It is convenient to rearrange the pairs as follows (I utilize $z_{4m-1} = z_0$, $z_{4m-1-u} = z_{-u}$, etc.):

$$z_0, z_u; z_u, z_{2u}; \cdots; z_{-u}, z_0.$$

This rearrangement will always be possible because the first elements of the pairs are the same for both sets, $0, u, 2u, \cdots$, running through the values $1, 2, 3, \cdots$ for $u = 1, 2, \cdots, 4m - 2$.

If u is a quadratic residue of $4m - 1$, then $u \equiv r^{2k}$ and we had seen that $z_{ir^{2k}} = z_i$. We may express the pairs $z_0, z_u; z_u, z_{2u}; z_{2u}, z_{3u}; \cdots; z_{-u}, z_0$ as $z_0, z_{r^{2k}}; z_{r^{2k}}, z_{2r^{2k}}; z_{2r^{2k}}, z_{3r^{2k}}; \cdots; z_{-r^{2k}}, z_0$; or finally as $z_0, z_1; z_1, z_2, z_2, z_3; \cdots, z_{-1}, z_0$. To summarize, $s_{1,u+1}$ is the number of pairs of adjacent elements z_j, z_{j+1} in a complete cycle $z_0, z_1, z_2, \cdots, z_{-1}, z_0$ for which $z_j \neq z_{j+1}$, if u is a quadratic residue of $4m - 1$.

If u is a quadratic nonresidue of $4m - 1$, then $u \equiv r^{2k-1}$ and we had seen that $z_{ir^{2k-1}} = z_{-1}$. In this case the pairs $z_0, z_u; z_u, z_{2u}; z_{2u}, z_{3u}; \cdots; z_{-u}, z_0$ may be expressed by $z_0, z_{r^{2k-1}}; z_{r^{2k-1}}, z_{2r^{2k-1}}; z_{2r^{2k-1}}, z_{3r^{2k-1}}; \cdots; z_{-r^{2k-1}}, z_0$; and finally by $z_0, z_{-1}; z_{-1}, z_{-2}; z_{-2}, z_{-3}; \cdots; z_1, z_0$. This time it is seen that $s_{1,u+1}$ is the number of pairs of adjacent elements z_j, z_{j-1} in a complete cycle $z_0, z_{-1}, z_{-2}, \cdots, z_1, z_0$ for which $z_j \neq z_{j-1}$, if u is a quadratic nonresidue of $4m - 1$.

But the two cycles, for u a residue and for u a nonresidue, give the same value of $s_{1,u+1}$, for one cycle is the other written backwards. It remains to find $s_{1,u+1} = s$, the distance $|a_j * a_k|$ for any distinct j, k. Consider the first digit of a_j, $j = 1, 2, \cdots, 4m - 1$. It will be z_j, which has been seen to take on the value 1 for $2m$ of the j and 0 for $2m - 1$ of the j. Exactly $2m(2m - 1)$ of the

$$\frac{(4m - 1)(4m - 2)}{2}$$

different $a_j * a_k$, $j \neq k$, will have 1 for the first digit. This is true also for the second, third, etc., digits by reason of the cyclic construction of the a_j. The total number of 1's in all the different $a_j * a_k$ is therefore $\sum_{j<k} |a_j * a_k| = (4m - 1)2m(2m - 1)$. But this sum is also

$$\frac{(4m - 1)(4m - 2)}{2}$$

because s was the distance between each pair and there are

$$\frac{(4m - 1)(4m - 2)}{2}$$

pairs a_j, a_k, $j \neq k$. Combining the two,

$$s = 2m = |a_j * a_k| \quad \text{for} \quad j \neq k;$$

$$j, k = k, 2, \cdots, 4m - 1.^6$$

Now that there have been constructed the a_j with the two desired properties $|a_j| = 2m$ and $|a_j * a_k| = 2m$ if $j \neq k$, it is easy to construct a code to demonstrate $A(4m, 2m) \geq 8m$.

[6] This implies that there are $2m$ alternations—1 followed by 0 or 0 followed by 1—in the sequence $z_0 z_1 z_2 \cdots z_{-1} z_0$. Since $z_1 = z_0 = 1$ and $z_{-1} = 0$, it follows that for primes of form $4m - 1$ the quadratic residues among $1, 2, \cdots, 4m - 2$ occur in exactly m blocks, as do the nonresidues.

For $j = 1, 2, \cdots, 4m - 1$, let b_j be the $4m$-digit point obtained by adding an 0 at the end of a_j. Because $|a_j| = 2m$ and $|a_j * a_k| = 2m$ for $j = k$ it is clear that $|b_j| = 2m$ and $|b_j * b_k| = 2m$ for $j = k$. I denote by e the $4m$-digit point whose digits are all 1; by o, as before, the $4m$-digit point whose digits are all 0.

I claim that the points e, o, b_j, $e * b_j$ form a code of $8m$ $4m$-digit points of minimum distance $2m$, demonstrating that $A(4m, 2m) \geq 8m$. Clearly there are $8m$ points in the code, each of $4m$ digits. Only the minimum distance requirement need be established. Since $|b_j| = 2m$ implies $|e * b_j| = 2m$, the zeros of b_j being the 1's of $e * b_j$ and conversely, it is clear that e, o are each at distance $2m$ from each of the remaining points. Also, $|b_j * b_k| = 2m$ for $j \neq k$ implies $|(e * b_j) * (e * b_k)| = 2m$ for $j \neq k$, the two distances being equal. It remains only to check $|b_j * (e * b_k)|$. But this is equal to $|e * (b_j * b_k)|$, and is equal to $2m$ or $4m$ because $|b_j * b_k| = 2m$ or 0 depending on whether $j \neq k$ or $j = k$.

This completes the proof that the code as constructed exemplifies $A(4m, 2m) \geq 8m$. The construction of such a code for given m is quite simple in practice, compared to the proof above that the code constructed fulfills the requirements. The case $m = 3$ is illustrated.

For $m = 3$, $4m - 1 = 11$. It is readily determined that 2 is a primitive root of 11, the numbers $2, 2^2, 2^3, \cdots, 2^{10}$ being congruent modulo 11 to 2, 4, 8, 5, 10, 9, 7, 3, 6, 1, respectively. The second, fourth, etc., of these are the residues: 4, 5, 9, 3, 1; the others 2, 8, 10, 7, 6, the non-residues. The definition of z_i requires that $z_i = 1$ for $i = 1, 3, 4, 5, 9$, and also for $i = 11$; $z_i = 0$ for $i = 2, 6, 7, 8, 10$. This gives us the a_j:

a_1: 1 0 1 1 1 0 0 0 1 0 1

a_2: 0 1 1 1 0 0 0 1 0 1 1

a_3: 1 1 1 0 0 0 1 0 1 1 0

. . . . . .

(etc., by cyclic permutation)

. . . . . .

a_{10}: 0 1 1 0 1 1 1 0 0 0 1

a_{11}: 1 1 0 1 1 1 0 0 0 1 0.

The desired code of $8m = 24$ points of $4m = 12$ digits each, of minimum distance $2m = 6$ is the following:

0:	000000 000000	e:	111111 111111
b_1:	101110 001010	$e * b_1$:	010001 110101
b_2:	011100 010110	$e * b_2$:	100011 101001
b_3:	111000 101100	$e * b_3$:	000111 010011
. . . . .		. . . . .	
. . . . .		. . . . .	
b_{10}:	011011 100010	$e * b_{10}$:	100100 011101
b_{11}:	110111 000100	$e * b_{11}$:	001000 111011.

On Decoding Linear Error-Correcting Codes—I*

NEAL ZIERLER†

Summary—A technique is described for finding simply computable numerical-valued functions of a received binary word whose value indicates where errors in transmission have occurred. Although it seems that a certain condition must usually be fulfilled for such functions to exist, or for our method to constitute an efficient procedure for finding them, there is, on the one hand, a strong tendency for "good" codes to satisfy the condition, while, on the other, it appears to be straightforward to construct codes which are good for a specified channel and also fulfill the condition. An advantage of the resulting decoding procedure is that it corrects and detects all possible errors; more precisely, if a word u is received and the coset $\bar{u}$ to which u belongs has a unique leader e, the procedure concludes that $u + e$ was sent, while if u has no unique leader, that fact, along with the weight of $\bar{u}$ (and sometimes a little more) can be indicated. The ideas and techniques are illustrated by the construction of decoding procedures for the perfect (23, 12) three-error-correcting code.

* Received by the PGIT, November 12, 1959. Operated with support from the U. S. Army, Navy and Air Force.

† Lincoln Lab., Mass. Inst. Tech., Lexington, Mass.

I. Introduction

THE type of decoding procedure with which this paper is concerned may be briefly described as follows (see also the Summary). Let V be the space of binary n-tuples; let A, the code, be a k-dimensional

ubspace and let L be an isomorphism of the space of l-cosets in V on the space S of binary n-k-tuples. Let ' denote a group of weight-preserving automorphisms f S (the weight of a coset is the smallest number of ones o be found among its members and L transfers this veight to S). Regard the members of S as rows and the nembers of Γ as matrices relative to the natural basis of . Γ decomposes S in two different ways into collections f subsets $\{0_i\}$ and $\{0_j'\}$ as follows: s and t are in the same $_i(0_j')$ if, and only if, $s\alpha = t$ ($s\alpha' = t$) for some $\alpha \in \Gamma$ where ' denotes the transpose of α. We shall see that the number of sets in these two decompositions is the same and that x is small, *i.e.*, approximately n, then an effective ecoding procedure consists essentially in determining o which 0_i a certain member of S, readily obtained as a unction of the received n-tuple, belongs. Furthermore, his determination may be made in the following way. et ψ_j, $j = 1, \cdots, x$ denote the numerical valued function on S defined as

$$\psi_j(s) = \sum_{t \in 0_j} \prod_{i=1}^{n-k} (-1)^{s_i t_i}.$$

Then the set of numbers $\psi_1(s), \cdots, \psi_x(s)$ are characteristic of the 0_i to which s belongs. Indeed, a properly hosen small number of the ψ_j will often suffice to distinguish the 0_i, for there is a tendency for all of the ψ_j o be rational functions of a few—and hence, the few nust separate the 0_i.

Our results are immediate descendents of some work of Prange [5], [6], [8] and were inspired in part by an observation of W. I. Wells.

The theory is developed in Sections II and III, while ections IV and V are devoted to an example—the construction of decoding procedures for the (23, 12) three-rror-correcting code of Golay-Paige—which is intended o illustrate the ideas and techniques in sufficient detail o serve provisionally as a model for further applications. The computations involved in designing the decoding rocedure are sometimes more easily carried out in V han S. The details of the way in which this may be done, ith an example, will be treated in a second paper.[1]

II. Preliminaries

Let $K = \{0, 1\}$ with addition and multiplication efined by $0 + 0 = 1 + 1 = 0$, $0 + 1 = 1 + 0 = 1$, $\cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0$, $1 \cdot 1 = 1$. Let n be a positive nteger and let $V_n = V$ be the vector space of n-tuples of lements of K; that is, an element of V is a $1 \times n$ matrix $= (u_1, \cdots, u_n)$ with coefficients (or "components") in he field K. Thus, V is a commutative group under the omponentwise, in K) addition of its elements, multiplication (componentwise) of members of V by elements of K is defined, and every subgroup of the group V is automatically a subspace of the vector space V.[2] For u, $v \in V$ and $B \subseteq V$, we define the *norm of* u, $||u||$, to be the number of nonzero components of u, the *distance from* u *to* v, $d(u, v)$, is $||u + v||$ and the *distance from* u *to* B, $d(u, B)$, is minimum $\{u + b: b \in B\}$. Assume $n > 1$ and let k be a positive integer less than n. An (n, k) *code* is determined by the choice of a k-dimensional subspace A of V and a linear transformation T of V_k on A. Two (n, k) codes with respective k-dimensional subspaces A and B of V are said to be *equivalent* if A is mapped on B by the automorphism $(u_1, \cdots, u_n) \to [u_{p(1)}, \cdots, u_{p(n)}]$ induced by some permutation $i \to p(i)$ of the indexes $1, \cdots, n$. It is easy to see ([7]) that any (n, k) code is equivalent to one with space A and linear transformation T of V_k on A such that $(vT')_i = v_i$ for $i = 1, \cdots, k$; we assume, henceforward, that A, T are a fixed k-dimensional subspace of V and linear transformation of V_k on A with this property. A *coset* of V modulo A is a subset of V of the form $\bar{v} = \{v + a: a \in A\}$ where v is some fixed element of V. Two cosets are either disjoint or identical, and the observation $u + v + a = u + v$ shows that the set V/A of cosets is a vector space over K under the addition $\bar{u} + \bar{v} = \overline{u + v}$. Since each coset contains 2^k members, V/A has $2^n/2^k = 2^{n-k}$ elements and so is isomorphic to V_{n-k}. For $u \in V$, we define the *weight of* u, $W(u)$ = the *weight of* $\bar{u}$, $W(\bar{u}) = d(0, \bar{u})$ = minimum $\{||v||: v \in \bar{u}\}$ and a *leader* of $\bar{u}$ is an element v of $\bar{u}$ with $||v|| = W(\bar{u})$. By a *decoding procedure* for a subset C of V (consisting of entire cosets) we shall mean a mapping D of C in V_k such that, for $u \in C$, $D(u)\ T + u$ depends only on the coset $\bar{u}$ to which u belongs and $||D(u)\ T + u|| = W(\bar{u})$ (see [7]). Clearly, D may be regarded as a function on part of V/A, and, once an isomorphism of V/A on V_{n-k} has been singled out, as a function on part of V_{n-k}.

Define the *inner product* $u \cdot v$ of two elements u and v of V by

$$u \cdot v = \sum_{i=1}^{n} u_i v_i$$

where the operations are those of K; if $u \cdot v = 0$, we say u is orthogonal to v, $u \perp v$; if $B \subseteq V$, $B^\perp = \{u \in V: u \perp b$ for all $b \in B\}$. Clearly, $A^\perp$ is a subspace of V of dimension $n - k$. Let L be an $n \times n - k$ matrix whose columns form a basis for $A^\perp$. Then uL depends only on the coset $\bar{u}$ to which u belongs, and $\bar{u} \to uL$ is an isomorphism of V/A on $S = V_{n-k}$; henceforward, we identify V/A and S under this isomorphism. A decoding procedure D of interest to us here has at its core a single real-valued function φ on V, which we call the "decoding function"

[1] Probably to be given at the Fourth London Symposium on nformation Theory.

[2] The chief use of the vector space viewpoint we shall make here is that endomorphisms of the groups involved may be represented by matrices with coefficients in K. For further details of the standard algebraic notions which appear here, see, *e.g.*, A. A. Albert, "Fundamental Concepts of Higher Algebra," University of Chicago Press, Chicago, Ill.; 1956.

of the procedure,[3] and the computation of a value of φ at a (suitably processed) received vector u yields, by relatively trivial computation, $D(u)$, or, perhaps, successively, the digits of $D(u)$. A consequence of the following lemma is that W may serve as a decoding function. Let $e_1, \cdots, e_n$ be the natural basis for V: $(e_i)_j = \delta_{ij}$; $i, j = 1, \cdots, n$.

Lemma 1:[4] Let $u \in V$, $1 \leq i \leq n$. Then $W(u + e_i) < W(u)$ if, and only if, some leader of $\bar{u}$ has a one in the ith place. Indeed, if $\bar{u}$ has a leader m with $m \cdot e_i = 1$ then $m + e_i$ is a leader of $u + e_i$, and if $W(u + e_i) < W(u)$ and v is a leader of $u + e_i$, then $v \cdot e_i = 0$ and $v + e_i$ is a leader of $\bar{u}$.

Proof: Suppose $W(u + e_i) < W(u)$ and let v be a leader of $u + e_i$. Then $W(u + e_i) = ||v|| < W(u) \leq ||v + e_i||$ since $v + e_i \in \bar{u}$. Hence, $||v + e_i|| = ||v|| + 1 = W(u)$ must hold, so $v \cdot e_i = 0$ and $v + e_i$ is a leader of $\bar{u}$. Conversely, suppose $\bar{u}$ has a leader m with $m \cdot e_i = 1$. Then, $W(u + e_i) = W(m + e_i) \leq ||m + e_i|| = ||m|| - 1 = W(u) - 1$, *i.e.*, $W(u + e_i) < W(u)$. It follows then by what we have just proved that if v is any leader of $u + e_i$, v has a 0 in the ith place and $v + e_i$, which has a one in the ith place, is a leader of $\bar{u}$. Hence, $||m + e_i|| = ||v||$ and $m + e_i$ is a leader of $u + e_i$.

Corollary 1: If $\bar{u}$ has the unique leader m and $m \cdot e_i = 1$, then $m + e_i$ is the unique leader of $u + e_i$.

Corollary 2: Let C be the subset of V consisting of those n-tuples whose cosets have unique leaders. Let $u \in C$, and define the sequence $u^0, u^1 \cdots, u^k$ as follows:

$$u^0 = u, \quad u^{i+1} = \begin{cases} u^i & \text{if } W(u^i + e_{i+1}) \geq W(u^i), \\ u^i + e_{i+1} & \text{if } W(u^i + e_{i+1}) < W(u^i) \end{cases}$$

$$i = 0, \cdots, k - 1.$$

Then the function D from C to V^k defined by $D(u) = u^k$ is a decoding procedure for C.

Corollary 3: If some leader of $\bar{u}$ has a one in the ith place, then every leader of $u + e_i$ has a 0 in the ith place.

We have as a final corollary a theorem of E. H. Moore and D. Slepian. The proof is a variation of the one discovered by E. Prange.

An n-tuple u is a *descendent* of an n-tuple v if $v \cdot e_i = 1$ whenever $u \cdot e_i = 1$; u is an *immediate descendent* of v if it is a descendent and $||u + v|| = 1$.

Define an ordering in V as follows: $u > v$ if, and only if, $u \neq v$ and if i is the smallest of the indexes j for which $u_j \neq v_j$, then $u_i = 1$.

Corollary 4: There exists a subset C of 2^{n-k} elements of V consisting exactly of one leader from each coset of V modulo A such that every descendent of an element of C belongs to C.

Proof: Let C be the set consisting of the largest leader of each coset (in the ordering of V defined above). It is clearly sufficient to verify that any immediate descendent of an element of C belongs to C. Let u be a member of C suppose $||u|| > 1$ (for otherwise there is nothing to prove) and suppose $u \cdot e_i = 1$. We must show that $u + e_i$ is the largest leader of $u + e_i$. Indeed, if v is any leader of $u + e_i$, then $v + e_i$ is a leader of $\bar{u}$ by the lemma, $v + e_i \leq u$ since $u \in C$. But $v \cdot e_i = 0$ by the lemma, $v \leq u + e_i$.

III. The Computation of Decoding Functions

Let V, A, L, S be as in Section II. For each $t \in S$ define the function X_t on S by $X_t(s) = (-1)^{st'}$ where t' denotes the transpose of the $1 \times n - k$ matrix t. The product $X_t X_w$ of two such functions is defined as a function on S by $X_t X_w(s) = X_t(s)\, X_w(s)$. It is readily verified that the set of all X_t: $t \in S$ forms a group $\mathcal{Y}$ under this multiplication which is isomorphic to S under $t \to X_t$. $\mathcal{Y}$ is called the *character group* of S and its elements are called the *characters of* S.

Let H be the set of all functions from S to the real numbers. H is evidently a real vector space of dimension 2^{n-k} with the natural inner product $f \cdot g = \sum_{s \in S} f(s)g(s)$ and corresponding norm: $||f|| = \sqrt{f \cdot f}$. Of course, each character is a member of H of norm $2^{(n-k)/2}$, and it is well-known and easy to see that $\mathcal{Y}$ is an orthogonal basis for H. Thus, if φ is a decoding function, we may design a device for computing $\varphi(u)$: $u \in V$ along the following lines. Assuming, as we do, that φ is constant on cosets, it may be identified with the function $f \in H$: $f(uL) = \varphi(u)$. We compute, once and for all, 2^{n-k} real coefficients c_t, indexed by the members of S, as follows:

$$c_t = \frac{1}{2^{n-k}} \sum_{s \in S} f(s) X_t(s).$$

Then

$$f(s) = \sum_{t \in S} c_t X_t(s) \tag{1}$$

provides a means of computing $\varphi(u)$, for we take $s = u$, and each $X_t(s)$ is a product of certain of the numbers $(-1)^{s_i}$. Of course, for the method to be practical, most of the coefficients c_t must be zero, and the nonzero ones must take on only a small number of different values. As will be seen, these things occur in a particularly convenient way if there exists a group Γ of automorphisms of S such that

$$f(s) = f(t) \text{ if, and only if, } t = s\alpha \text{ for some } \alpha \in \Gamma. \tag{2}$$

Define an *orbit* of a given group Γ of automorphisms of S to be a subset of S of the form $s\Gamma = \{s\alpha: \alpha \in \Gamma\}$ for some $s \in S$ and define a *level set* of a function g on S to be the complete subset of S on which g takes on a given value. Evidently, distinct orbits are disjoint. Condition (2) may be restated as follows: the orbits of Γ coincide with the level sets of f. Suppose now that we begin the design of our decoding procedure by choosing a tentative decoding function g, a function on S, for which there exists a non-trivial group Γ of automorphisms of S under which g is invariant in the sense that $g \cdot \alpha = g$ for all $\alpha \in \Gamma$ [wh

[3] Since φ takes on only a finite set of values, it may, of course, be restricted to be integer valued.

[4] This is a slight refinement of a discovery of E. Prange.

$\cdot \alpha(s) = g(s\alpha)]$, *i.e.*, g is constant on the orbits of Γ. We hall exhibit a natural "computable" orthogonal basis for he functions on S constant on the orbits of Γ in terms of hich a decoding function f may be constructed which ends to achieve the desired reduction of (1). We suppose he elements of Γ to be $n - k$ square matrices (*i.e.*, we epresent the linear transformations of Γ by matrices elative to the natural basis of S) and let α' denote the ranspose of the matrix α.

Lemma 2: $X_t \cdot \alpha = X_{t\alpha'}$.

Proof: $X_t \cdot \alpha(s) = X_t(s\alpha) = (-1)^{s\alpha t'}$

$= (-1)^{s(t\alpha')'} = X_{t\alpha'}(s)$.

Let $\Gamma' = \{\alpha' : \alpha \in \Gamma\}$; clearly Γ' is a group of $n \times n$ atrices isomorphic to the group Γ under $\alpha' \to \alpha^{-1}$. et $0_1', \cdots, 0_{x'}'$ be the distinct orbits of S under Γ' and t ψ_i be the integer-valued function on S defined as ollows:

$$\psi_i(s) = \sum_{t \in 0_i'} X_t(s), \qquad i = 1, \cdots, x'.$$

et H_Γ denote the set of all real-valued functions on S hich are constant on the orbits $0_1, \cdots, 0_x$ of S under Γ; early, H_Γ is a real vector space of dimension x.

Theorem 1: The functions $\psi_1, \cdots, \psi_{x'}$ form an orthogonal asis for H_Γ. In particular, $x' = x$.

Proof: Since $\mathcal{Y}$ is a set of pairwise orthogonal functions n S, the linear spans of two disjoint subsets of $\mathcal{Y}$ are thogonal and, hence, the ψ_i are pairwise orthogonal. If $\in S$ and $\alpha \in \Gamma$,

$$\psi_i(s\alpha) = \sum_{t \in 0_i'} X_t(s\alpha) = \sum_{t \in 0_i'} X_{t\alpha'}(s)$$

$$= \sum_{t \in 0_i' \alpha'} X_t(s) = \sum_{t \in 0_i'} X_t(s) = \psi_i(s),$$

., the ψ_i belong to H_Γ. It follows that $x' \leq x$, and inter-anging the roles of Γ and Γ' (which we may do since $'' = \Gamma$), $x \leq x'$. Hence, $x = x'$ and the result follows.

Of course, the significant thing about a decoding func-on is not the precise values it takes on, but rather the ecomposition of S into level sets which it induces. By heorem 1 we can choose a set (in a suitable sense mini-al) of ψ_i which separates the level sets of a tentative ecoding function g (with $g \in H_\Gamma$), and then some simple near combination f of these ψ_i will have the same level ts as g, or some still simpler linear combination of the may serve as a new decoding function. It should be noted at the desideratum for Γ is not that it be large, but that be transitive, or nearly so, on each level set of g. That given s and t in the same level set, we should like Γ to ntain an α with $s\alpha = t$; but a large Γ which thoroughly ixes the elements of small subsets of level sets might ll far short of transitivity.

Most of the basic ideas will, hopefully, be clarified by e discussion of the example which follows. This section concluded with a few further remarks on the weight nction.

An automorphism of a space V_i is, of course, determined by the assignment of an image to each member of a basis; an automorphism which maps the natural basis on itself is called a *coordinate permutation*. In general, the statement that an automorphism M of V_i leaves the subset C of V_i "invariant" or "maps it on itself" means that $cM \in C$ for every $c \in C$ (not necessarily that $cM = c$ for every $c \in C$, *i.e.*, that C is *pointwise invariant*).

In view of Lemma 1, the weight function W appears as a natural candidate for the decoding function. The following theorem gives a characterization of the full group of automorphisms of S which preserve W.

Theorem 2: Suppose $||a|| > 2$ for every $a \in A$ except 0. Let π be the group of coordinate permutations of V which leave A invariant and let Γ be the group of automorphisms of S which preserve W. An isomorphism $P \to M_P$ of π on Γ is established as follows: $uLM_P = uPL$.

Proof: Let $P \in \pi$ and let $M = M_P$. If $uL = vL$ then $u + v \in A$ so $(u + v)P \in A$ and $0 = (u + v)PL = uPL + vPL$, *i.e.*, $uPL = vPL$, and it follows that M is well-defined as a function from S to S. If u and v are arbitrary members of V, $(uL + vL)M = (u + v)LM = (u + v)PL = uPL + vPL = uLM + vLM$ and M is an endomorphism of S. If $uLM = 0$ then $uPL = 0$ so $uP \in A$; hence, $u \in A$, so $uL = 0$, *i.e.*, M is an automorphism of S. Since P obviously preserves weight, we have $W(uLM) = W(uPL) = W(uP) = W(u)$ and $M \in \Gamma$. If $P, Q \in \pi$, $uLM_{PQ} = uPQL = (uP)QL = (uP)LM_Q = (uPL)M_Q = uLM_PM_Q$, and $P \to M_P$ is a homomorphism of π in Γ. If $M_P = 1$, then $uPL = uLM_P = uL$ so $uP + u \in A$ for all $u \in V$. In particular, $d_i = e_iP + e_i \in A$, and since $||d_i|| = 0$ or 2, $e_iP = e_i$ must hold by hypothesis and hence $P = 1$, *i.e.*, $P \to M_P$ is an isomorphism of π in Γ. It remains to show that the mapping is onto Γ. To do this, we choose $M \in \Gamma$ arbitrarily and construct $P \in \pi$ such that $M = M_P$. Indeed, $S_1 = \{e_iL\}_{i=1}^n$ are the elements of S of weight 1, and since S_1 is invariant under M by assumption, there exists a mapping θ of $\{1, \cdots, n\}$ in itself such that $e_iLM = e_{\theta(i)}L$. If $\theta(i) = \theta(j)$ then $0 = e_{\theta(i)}L + e_{\theta(j)}L = e_iLM + e_jLM = (e_i + e_j)LM = (e_i + e_j)L$ since M is nonsingular so $e_i + e_j \in A$ and hence $i = j$, since A contains no elements of norm 2. Thus, θ is a permutation and we let P denote the coordinate permutation with $e_iP = e_{\theta(i)}$. If $u \in V$, $uPL = \sum u_ie_iPL = \sum u_ie_{\theta(i)}L = \sum u_ie_iLM = uLM$; hence, if $u \in A$, $uL = 0$ implies $uPL = 0$ implies $uP \in A$ so $P \in \pi$ and the same formula shows that $M = M_P$, thereby completing the proof of Theorem 2.

Corollary 5: Suppose $||a|| > 2$ for $a \neq 0 \in A$ and let M be an automorphism of S. Then M preserves weights if, and only if, S_1, the set of elements of S of weight 1, is invariant under M.

Corollary 6: Suppose $||a|| > 2$ for $a \neq 0 \in A$ and let M be an automorphism of S. Then M preserves weights if, and only if, it permutes the rows of L.

Corollary 7: With the assumption and notation of the statement, let α be any automorphism of V leaving A invariant and preserving W. Then the permutation of

cosets induced by α coincides with that induced by some member of π.

IV. Decoding Procedures for the (23, 12) Three-Error-Correcting Code of Golay-Paige[5]

For this code, $n = 23$, $k = 12$, and, following [4], we choose $L =$

$$\begin{array}{cccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1
\end{array}.$$

Observe that A, which is, of course, uniquely determined by L, has the property that for each $v \in V_k$ there is (a necessarily unique) $a \in A$ with $v_1 = a_1, \cdots, v_k = a_k$, and it is natural to take for T this correspondence.[6] The key feature of the code is that every coset has a unique leader of norm ≤ 3, and every element of V of norm ≤ 3 is a coset leader.

Consider the following two permutations of the natural basis of S:

$$\alpha = (1, 4, 2)\ (3, 5, 6)\ (8, 10,9)\ (7)\ (11),$$

$$\gamma = (1, 2, 3)\ (4, 7, 5, 8, 6, 9)\ (10,11).$$

The notation is as follows: α is the permutation which leaves e_7 and e_{11} fixed, maps $e_1 \to e_4 \to e_2 \to e_1$, etc. If α^*, γ^* are the corresponding induced automorphisms of S, $\alpha \to \alpha^*$ and $\gamma \to \gamma^*$ establishes an isomorphism of the group Γ of permutations of the set $[1, \cdots, 11\}$ generated by α and γ on the group Γ^* of automorphisms of S generated by α^* and γ^*. It is easily checked that Γ^* permutes the rows of L and so, by Corollary 6 of Theorem 2 (which applies since every $a \neq 0 \in A$ has norm ≥ 7 W is invariant under Γ^*. More specifically, if we numbe the rows of L $r_1 - r_{23}$ (and regard them as membe of S), the following are invariants of Γ^*:

1) $\{r_1, \cdots, r_{11}\}$,
2) $\{r_{12}\}$,
3) $\{r_{13}, \cdots, r_{23}\}$ and
4) $||s||$ for all $s \in S$.

Indeed 1), 2) and 4) are immediate since Γ^* consists coordinate permutations and 3) is readily checked for α and γ^* (and, of course, accounts for their choice). Let u classify the coset leaders or "errors" into "types" wit the following notation: a leader u is of type $(i, j, m$ where each i, j, n is a nonnegative integer, $0 \leq i + j +$ ≤ 3, if u has i ones in its first 11 places, $u_{12} = j$, and ones in the last 11 places. Thus, 1) to 3) may be para phrased as follows: coset leader type is invariant unde Γ^*. The foregoing suggest that we take for decoding fun tion the refinement φ of W: $\varphi(u) = (i, j, m)$ where u is vector whose coset leader is of type (i, j, m) [in th notation, $W(u) = i + j + m$]. Of course, L provides natural $1 - 1$ mapping of coset leaders onto S, so tha we may speak of the "type" of $s \in S$; let f be the functio on S: $f(uL) = (||uL||, i, j, m)$ where $\varphi(u) = (i, j, m)$ The following theorem will be established in the ne section.

Theorem 3: The orbits of Γ^* are distinct level sets of with the following exception: the set on which f takes the value (3, 3, 0, 0) consists of two orbits of Γ^*.

The function f appears to be a little finer tha necessary—for the most part we don't care about $||uL||$- but the idea is that Γ^* provides a convenient method computing f, from which W is trivially obtained, an the additional information with regard to error type ma be used to make the decoder more efficient. It is co venient to introduce what amounts to another notatio for the members of S. Let S^* be the set of subsets $I = \{1, \cdots, 11\}$ and for $s \in S$ let $s^* \in S^*$: $i \in s^*$ if, a only if, $s_i = 1$. The elements of Γ act as permutations S^* in the natural way: *e.g.*, if $s^* = \{1, 2, 7\}$ then $s^*\alpha$ $\{4, 1, 7\}$ (*i.e.*, α transforms each element of s^*). Clearl for $\beta \in \Gamma$ and $s \in S$, $(s\beta^*)^* = s^*\beta$; *i.e.*, roughly speakin replacing S by S^* replaces Γ^* by Γ.

Note that for $\beta^* \in \Gamma^*$, $\beta^{*\prime} = \beta^{*-1}$, since β^* is a coordina permutation and, hence, $\Gamma^{*\prime} = \Gamma^*$ (in other words, taki transposes in Γ^* corresponds to taking inverses in It will be shown in Section V that Γ is doubly transitiv hence, the elements of S of norm 2 fall into a single orb as do those of norm 1. Let the function ψ correspondi to the latter orbit be denoted ψ_1; thus, $\psi_1(s) = 11 - 2$ [and the ψ function corresponding to the former orbit $(\psi_1^2 - 11)/2$].

[5] For an equivalent cyclic code with decoding procedure, see [5, 8].
[6] See Section 4 of [4].

[7] f is a refinement of $\varphi \cdot L$; of course, it is not integer or real-val but it could easily be made so; it seems more natural to treat it its present form.

There are altogether 24 orbits,[8] which occur naturally in pairs, for if θ is an orbit in S^*, so is $\bar{\theta} = \{\bar{s^*}: s^* \in \theta\}$ where $\bar{s^*} = \{i \in I: i \notin s^*\}$. Clearly, $||\bar{s^*}|| = 11 - ||s^*||$. As already noted, the elements of norm 2 fall into a single orbit, as do those of norm 1 (and, of course, $0 \in S \equiv \phi \in S^*$ is an orbit by itself). There are two orbits of norm 3, three of norm 4 and four of norm 5. The orbits of norms 0, 1, 2 are named 0, 1, 2, respectively; orbits of norm 3–5 are named in the form $i \cdot j$ where i is the norm and j is an index = 1, $\cdots$ ("numerical order" by "smallest" member). An orbit of norm > 5 is the complement of some orbit of norm ≤ 5 and is given its name primed.

TABLE I
ORBITS 3.1 AND 5.3

5.3		3.1			5.3		3.1		
10	11	1	2	3	8	10	3	4	5
7	8	1	2	5	1	6	3	4	11
4	6	1	2	9	7	11	3	5	6
8	9	1	3	6	1	2	3	5	9
4	5	1	3	7	2	4	3	6	10
9	11	1	4	5	6	10	3	7	9
7	10	1	4	6	2	8	3	7	11
2	3	1	4	8	4	11	3	8	9
3	6	1	5	10	1	7	3	8	10
2	5	1	6	11	5	9	3	10	11
6	11	1	7	8	2	6	4	5	7
2	9	1	7	10	3	5	4	6	9
5	10	1	8	9	1	8	4	7	9
3	7	1	9	11	3	11	4	7	10
4	8	1	10	11	6	9	4	8	10
7	9	2	3	4	5	7	4	8	11
5	6	2	3	8	2	10	4	9	11
8	11	2	4	6	1	4	5	6	8
1	5	2	4	10	3	9	5	7	8
9	10	2	5	6	1	10	5	7	11
3	4	2	5	11	2	11	5	8	10
1	3	2	6	7	4	7	5	9	10
4	10	2	7	8	6	8	5	9	11
5	11	2	7	9	5	8	6	7	10
1	9	2	8	11	4	9	6	7	11
3	8	2	9	10	2	7	6	8	9
6	7	2	10	11	3	10	6	8	11
					1	11	6	9	10

Table I lists simultaneously an orbit of norm 3 and one of norm 5 (together with the trivial one of norm 2), both of which are of length 55.

Orbit 5.4 is of length 11 and its members are the rows in the following array:

1	2	4	7	11
1	2	6	8	10
1	3	4	9	10
1	3	5	8	11
1	5	6	7	9
2	3	5	7	10
2	3	6	9	11
2	4	5	8	9
3	4	6	7	8
4	5	6	10	11
7	8	9	10	11 .

[8] See the following section for proofs and further details.

Let ψ_2 and ψ_3 correspond to orbits 3.1 and 5.4, respectively. Thus, letting

$$s_i' = 1 - 2s_i, \qquad \psi_2(s) = s_1's_2's_3' + s_1's_2's_5' + s_1's_2's_9' + \cdots + s_6's_9's_{10}'$$

and

$$\psi_3(s) = s_1's_2's_4's_7's_{11}' + \cdots + s_7's_8's_9's_{10}'s_{11}';$$

note, however, that combinatorial properties of the orbits lead to potentially much more efficient means of computing the ψ's—see Section V.

TABLE II

Orbit	Representative	Error Type	ψ_1	ψ_2	ψ_3	Orbit Length
0	ϕ	0,0,0	11	+55	+11	1
1	1	1,0,0	9	+25	+1	11
2	1,2	2,0,0	7	+7	−1	55
3.1	1,2,3	3,0,0	5	−7	+5	55
3.2	1,2,4	3,0,0	5	−11	−3	110
4.1	1,2,3,4	2,0,1	3	−9	+3	
4.2	1,2,3,10	0,0,3	3	−1	−5	
4.3	1,2,4,7	1,1,1	3	+7	+3	
5.1	1,2,3,4,5	1,0,1	1	−15	+1	
5.2	1,2,3,4,7	1,0,2	1	+2	+1	
5.3	1,2,3,4,8	0,1,2	1	−7	−7	55
5.4	1,2,4,7,11	0,1,1	1	+25	+9	11
5.4′		0,0,1	−1	−25	−9	11
5.3′		0,0,2	−1	+7	+7	55
5.2′		2,0,1	−1	−2	−1	
5.1′		1,1,1	−1	+15	−1	
4.3′		1,0,1	−3	−7	−3	
4.2′		1,0,2	−3	+1	+5	
4.1′		1,0,2	−3	+9	−3	
3.2′		2,0,1	−5	+11	+3	110
3.1′		0,0,3	−5	+7	−5	55
2′		2,1,0	−7	−7	+1	55
1′		1,1,0	−9	−25	−1	11
0′		0,1,0	−11	−55	−11	1

Table II lists a representative of each orbit in S^* of norm ≤ 5 (its "numerically smallest" member) together with the values of φ (the error = coset leader type), ψ_1, ψ_2 and ψ_3, and a few orbit lengths. Of course, in general, $\psi(s^*) = -\psi(s^*)$.

The level sets of the function $g = (\psi_1, \psi_2)$ [the value $g(s)$ of a vector s is the ordered pair of numbers $\psi_1(s)$,

$\psi_2(s)]$ are exactly the orbits of Γ. It is a slight refinement of f [it distinguishes the two orbits of type (3, 0, 0)] and appears to be an acceptable candidate for the role of decoding function.

The pair $h = (\psi_1, \psi_3)$ does not quite separate the level sets of f (nor those of W) but in view of the relative simplicity of the function ψ_3 it seems worthwhile to sketch a decoding procedure in which h is the decoding function. The motivating notion here is that if we find a function that falls a little short of accomplishing the desired separation of level sets (but is otherwise advantageous, *e.g.*, simple to compute) then it may be possible, in effect, to use it as a decoding function by bringing into play some further properties of the code.

Let $d_1, \cdots, d_{23}$ and $e_1, \cdots, e_{11}$ be the natural bases in V and S, respectively. We assume that some $v \in V_{12}$ has been chosen as message, that $vT \in A$ has been sent and $u \in V$ has been received with $||vT + u|| \leq 3$. We shall have occasion to consider, along with $uL = s$,

$$(u + d_i)L = uL + d_iL = s + e_i, \qquad i = 1, \cdots, 11.$$

Let

$$\psi_i^{(j)} = \psi_i(s + e_j), \quad h_j = [\psi_1^{(j)}, \psi_3^{(j)}], \quad j = 1, \cdots, 11.$$

The decoding operation might proceed along the following lines.

Compute $\psi_1(s)$.

1) If $\psi_1(s) = 11$, $v_i = u_i$, $i = 1, \cdots, 12$.
2) If $\psi_1(s) = 9$, 7 or 5 there are 1, 2 or 3 errors respectively among the first 11 u_i (and none elsewhere). Compute $\psi_1^{(i)}(s)$. If $\leq \psi_1(s)$, $v_i = u_i$; if $> \psi_1(s)$ (in fact, $= \psi_1(s) + 2$), $v_i = u_i + 1$. Terminate by finding the right number of errors or by replacing s by $s + e_i$ if i is smallest integer for which $\psi_1^{(i)}(s) > \psi_1(s)$, etc., and terminate when $\psi_1 = 11$ (see Corollary 2 of Lemma 1).
3) If $\psi_1(s) = -11$, $v_i = u_i$, $i = 1, \cdots, 11$, $v_{12} = u_{12} + 1$.
4) If $\psi_1(s) = -7$ or -9 there is an error in u_{12} and 2 of 1, respectively, in the first 11 places. Complement s, set $v_{12} = u_{12} + 1$ and go to 2 for $v_1, \cdots, v_{11}$. If $\psi_1(s) \notin \{11, 9, 7, 5, -7, -9, -11\}$, compute $\psi_3(s)$, thus obtaining $h(s) = [\psi_1(s), \psi_3(s)]$.
5) If $h(s) = (-5, -5)$, $(-1, 7)$, $(-1, -9)$ or $(3, -5)$, there are no errors in first 12 places, *i.e.*, $v_i = u_i$, $i = 1, \cdots, 12$.
6) If $h(s) = (1, -7)$ or $(1, 9)$, $v_i = u_i$, $i = 1, \cdots, 11$, $v_{12} = u_{12} + 1$.
7) $h(s) = (3, 3)$ or $(-1, -1)$. As a glance at Table II will show, this is the first ambiguity, for the error in u may be (in either case) of either type (2, 0, 1) or (1, 1, 1). Here, however, we can make use of further properties of A as follows. Start computing $h_i(s)$, $i = 1, 2, \cdots, 11$. We obtain eventually the value (1, 9) if *and only if* $\varphi(u) = (1, 1, 1)$. In this case, if i is the place where the value was obtained, $v_i = u_i$ for $12 \neq j \neq i$, $v_i = u_i + 1$ for $j = 12, i$. If (1, 9) is never obtained, $\varphi(u) = (2, 0, 1)$ and we proceed as follows [of course, this could be done along with the search for (1, 9)]. Let i be the first place in which $h_i(s) = (1, 1)$ or $(-3, -3)$. Then compute $h_j(s + e_i)$, $i \neq j = 1, \cdots, 11$. If, for some j, $h_j(s + e_i) = (-1, -9)$, then the two errors in the first 11 places of u are in places i and j. If $h_j(s + e_i) = (-1, -9)$ holds for no j, take for i and *second* value for which $h_i(s) = (1, 1)$ or $(-3, -3)$ and repeat.
8) $h(s) = (1, 1)$ or $(-3, -3)$. Then for some $i = 1, \cdots, 11$, $h_i(s) = (-1, -9)$ respectively $(-1, 7)$ if *and only if* $\varphi(u) = (1, 0, 1)$ respectively $(1, 0, 2)$, and for this i, $v_j = u_j$, $j \neq i$, $v_i = u_i + 1$.
9) $h(s) = (-3, 5)$. Choose i: $h_i(s) = (-1, 7)$. Then $v_i = u_i + 1$, $v_j = u_j$, $i \neq j = 1, \cdots, 12$.
10) $h(s) = (-5, 3)$. Then $\varphi(u) = (2, 0, 1)$ and we search for i with $h_i(s) = (-3, -3)$; then compute $h_j(s + e_i)$, $i \neq j = 1, \cdots, 11$. If $h_j(s + e_i) = (-1, -9)$ for some j, the two errors in the first eleven places of u are in the ith and jth places. If, however, $h_j(s + e_i) = (-1, -9)$ holds for no j, take for i the second value for which $h_i(s) = (-3, -3)$ and repeat.

V. Data, etc.

This section gathers together proofs and details omitted from Section IV.

The Group Γ.

Γ is a group of permutations of $I = \{1, \cdots, 11\}$ with two generators α and γ; we shall show that it is a doubly transitive group of order 660. First, we list some useful elements of Γ.

$$\begin{aligned}
\alpha &= (1, 4, 2)\,(3, 5, 6)\,(8, 10, 9),\\
\gamma &= (1, 2, 3)\,(4, 7, 5, 8, 6, 9)\,(10, 11),\\
\gamma^2 &= (1, 3, 2)\,(4, 5, 6)\,(7, 8, 9),\\
\alpha^2\gamma\alpha &= (1, 5, 4)\,(2, 7, 6, 10, 3, 8)\,(9, 11),\\
\alpha\gamma^2\alpha^2 &= (2, 6, 4)\,(1, 3, 5)\,(7, 9, 10),\\
\alpha\gamma &= (1, 7, 5, 9, 6)\,(3, 8, 11, 10, 4),\\
&(\beta \text{ and } \delta \text{ are defined below})\\
\delta\beta^2\gamma\delta\beta^2 &= (1, 2, 5)\,(6, 10, 3, 11, 4, 9)\,(7, 8),\\
\gamma\delta &= (1, 2, 3)\,(4, 6, 5)\,(7, 9, 8),\\
\gamma\alpha &= (2, 5, 10, 11, 9)\,(3, 4, 7, 6, 8),\\
\alpha_1 &= \alpha\gamma^2\alpha\gamma^2\alpha^2\\
&= (2, 5)\,(3, 4)\,(7, 8)\,(9, 10),\\
\beta_1 &= (\alpha^2\gamma\alpha)\,(\alpha\gamma^3\\
&= (2, 6, 8)\,(3, 4, 9)\,(5, 11, 7),
\end{aligned}$$

} 1 fixed

$$\begin{aligned}
\delta &= \gamma\alpha^2\gamma^2\alpha\gamma^2\alpha^2\\
&= (4, 8)\,(5, 9)\,(6, 7)\,(10, 11),\\
\beta &= (\gamma\alpha)^2\alpha^2\gamma\alpha\,(\alpha\gamma)^3\,(\gamma\alpha)^3\\
&= (4, 7, 11)\,(6, 8, 10)\,(3, 9, 5),\\
\delta\beta &= (3, 9)\,(4, 10)\,(6, 11)\,(7, 8),\\
\delta\beta^2 &= (3, 5)\,(4, 6)\,(8, 11)\,(7, 10).
\end{aligned}$$

} 1 and 2 fixed

Rows 13–23 of L, regarded as members of S^* and labelled $a = k$ (in a different order) are:

$a = \{1, 2, 3, 4, 5, 6\}$,
$b = \{1, 2, 3, 7, 8, 9\}$,
$c = \{1, 2, 5, 9, 10, 11\}$,
$d = \{1, 3, 6, 7, 10, 11\}$,
$e = \{1, 4, 5, 7, 8, 10\}$,
$f = \{1, 4, 6, 8, 9, 11\}$,
$g = \{2, 3, 4, 8, 10, 11\}$,
$h = \{2, 4, 6, 7, 9, 10\}$,
$i = \{2, 5, 6, 7, 8, 11\}$,
$j = \{3, 4, 5, 7, 9, 11\}$,
$k = \{3, 5, 6, 8, 9, 10\}$.

Let Γ_{12} be the six-element subgroup of Γ consisting of δ, β, β^2, $\delta\beta$, $\delta\beta^2$ and the identity; evidently, the numbers 1 and 2 are both fixed by all the members of Γ_{12}. We shall show that, in fact, Γ_{12} is the full subgroup of Γ fixing both 1 and 2. Indeed, suppose $\eta \in \Gamma$ fixes both 1 and 2. Observe first that Γ is doubly transitive, for 1 is mapped into 2, 3, $\cdots$, 11, respectively, by α^2, γ^2, α, $\alpha\gamma^2$, $\alpha\gamma^2\alpha$, $\alpha\gamma$, $\alpha\gamma^3$, $\alpha\gamma^5$, $\alpha\gamma^3\alpha$, $\alpha\gamma^3\alpha\gamma$, respectively, proving that Γ is transitive. In the subgroup fixing 1, 2 is mapped on 3, 4, $\cdots$, 11, respectively, by $(\gamma\alpha)^4\beta_1$, $(\gamma\alpha)^4\beta_1^2$, α_1, β_1, $\alpha_1\beta_1^2$, β_1^2, $(\gamma\alpha)^4$, $(\gamma\alpha)^2$, $(\gamma\alpha)^3$, respectively, proving that Γ is doubly transitive. Note, however, that Γ is not triply transitive, for if it were, there would be an element ε fixing 1 and 2 and mapping 3 on 4. But then $a\varepsilon = a$ must hold, since $\{1, 2, 3\} \subseteq a$ implies $\{1, 2, 4\} \subseteq a\varepsilon$, and a alone contains 1, 2 and 4. But similarly, $b\varepsilon = a$ must hold, and this contradiction shows that Γ is not triply transitive. Returning now to the arbitrary element η of Γ fixing 1 and 2, if $3\eta \notin \{3, 5, 9\}$, a glance at Γ_{12} shows that the subgroup of Γ fixing 1 and 2 is transitive, hence that Γ is triply transitive, which is impossible. Hence, $3\eta \in \{3, 5, 9\}$, so the product of η and some element of Γ_{12} fixes 3; if we show that this product is in Γ_{12}, it will follow that η is in Γ_{12}. In other words, it is sufficient to show that $3\eta = 3$ implies $\eta = \delta$ or the identity. Now since η fixes 1, 2 and 3, it has the following invariants: $\{a, b\}$, $\{e, f\}$, $\{j, k\}$, $\{c\}$, $\{d\}$ and $\{g\}$. Suppose $a\eta = b$. Then $\{4, 5, 6\} \to \{7, 8, 9\} \to \{4, 5, 6\}$ and $g\eta = g$ implies $4\eta = 8$, $8\eta = 4$. Similarly, $d\eta = d$, $c\eta = c$, respectively, imply that η contains the cycles (6, 7) and (5, 9) respectively. The foregoing implies $e\eta = f$ which in turn implies that $10\eta = 11$, which completes the proof that $\eta = \delta$ on the assumption that $a\eta = b$. The remaining possibility is that η fixes a and b, as well as c, d and g. Hence, the following are invariants of η: $\{4, 5, 6\}$, $\{7, 8, 9\}$, $\{5, 9, 10, 11\}$, $\{6, 7, 10, 11\}$ and $\{4, 8, 10, 11\}$; comparison of these shows at once that η fixes, in addition to 1, 2 and 3, the numbers 4, $\cdots$, 9. It follows that $e\eta = e$ must hold, and hence, since $10\eta = 10$, that η is the identity. Thus, Γ_{12} is the full subgroup of Γ fixing 1 and 2 we have[9] the following result.

Theorem 4: Γ is doubly transitive and its order is $11 \cdot 10 \cdot 6$. Furthermore, any maximal subgroup of Γ fixing two numbers is isomorphic to the symmetric group of degree 3.

The Orbits of Γ

Since ψ_1, ψ_2 separate the representatives of orbits listed in Table II, the representatives do indeed lie in distinct orbits. We shall show that there are no others. It is sufficient, of course, to consider orbits of norm ≤ 5, and those of norm ≤ 2 are disposed of at once by double transitivity. Now every orbit of norm > 2 contains members containing 1 and 2 by double transitivity. We choose as representative for each orbit its numerically smallest member[10] which will then automatically begin 1, 2, $\cdots$. Let us begin by considering orbits of norm 4 and suppose that $\{1, 2, x, y\}$ is the representative of an orbit. If $x \neq 3$, then $x, y \neq 5, 9$, for otherwise we could apply a member of Γ_{12} to replace x or y by 3, thereby obtaining a numerically smaller member of the same orbit. Thus, $\{x, y\} \subseteq \{4, 6, 7, 8, 10, 11\}$ which implies $x = 4$, $y \in \{6, 7, 8, 10, 11\}$. Then $y \neq 6$, for $\{1, 2, 4, 6\}$ $\alpha = \{1, 2, 4, 9\}$. Thus, $y = 7$ or 11, and since $\{1, 2, 4, 7\}$ $\gamma^3\delta\beta^2 = \{1, 2, 4, 11\}$, $y = 7$ must hold. We have proved that there is only one orbit of norm 4 that does not have representative of the form $\{1, 2, 3, y\}$ and that its representative is $\{1, 2, 4, 7\}$. Now consider representatives of the form $\{1, 2, 3, y\}$. A glance at γ shows that $y = 4$ or $y = 10$ must hold, and these do, indeed, yield distinct orbits.

Consider now a representative of an orbit of norm 5 of the form 1, 2, 3, 4, x. If $x \neq 5$ then $x \neq 6$ for $\{1, 2, 3, 4, 6\}$ $\alpha = \{1, 2, 3, 4, 5\}$. Further, $\{1, 2, 3, 4, 11\}\alpha\delta\beta^2\gamma^3 = \{1, 2, 3, 4, 7\}$, $\{1, 2, 3, 4, 9\}$ $\gamma = \{1, 2, 3, 4, 7\}$ and $\{1, 2, 3, 4, 10\}$ $\beta^2\alpha\gamma^5\alpha^2\gamma^3\delta\beta^2\gamma^5\delta\beta^2\alpha\gamma^2\alpha^2\gamma = \{1, 2, 3, 4, 7\}$. It follows that if a representative is of the form $\{1, 2, 3, 4, x\}$ then $x = 5$, 7 or 8 (and all of these occur). Now consider a representative $\{1, 2, 3, x, y\}$ with $4 < x < y$; a glance at γ shows that the only possibility for x and y is $x = 10$, $y = 11$; but $\{1, 2, 3, 10, 11\}$ $\delta\beta^2\gamma^5\alpha_1 = \{1, 2, 3, 4, 8\}$. It remains to show that there is just one orbit with representative $\{1, 2, x, y, z\}$ with $3 < x$. There is indeed one, for a, b, $\cdots$, j, k is an orbit, and so is its complement which is orbit 5.4 with representative $\{1, 2, 4, 7, 11\}$. Now, assuming $3 < x < y < z$, $\{x, y, z\} \cap \{5, 9\} = \phi$, for otherwise we could apply a member of Γ_{12}, which is transitive in $\{3, 5, 9\}$, to obtain a member of the orbit of the form $\{1, 2, 3, y', z'\}$. Thus, $\{x, y, z\} \subseteq \{4, 6, 7, 8.$

[9] See Burnside [1], Section 137, Theorem III.

[10] Let $E = \{n_1, \cdots, n_s\}$, $F = \{m_1, \cdots, m_s\}$ be subsets of I, each with s members, and suppose $n_1 < \cdots < n_s$ and $m_1 < \cdots < m_s$. "E is numerically smaller than F" if when t is the smallest index such that $n_t \neq m_t$, then $n_t < m_t$.

10, 11}, and since Γ_{12} is transitive in this set, we must have $x = 4$ and $\{y, z\} \subseteq \{6, 7, 8, 10, 11\}$. Now 8 and 10 must be excluded, for, *e.g.*, $\{1, 2, 4, 8, z\}\ \alpha = \{1, 2, 4, 9, z\alpha\}$, $\{1, 2, 4, 9, z\alpha\}\ \delta\beta = \{1, 2, 3, 10, z\alpha\delta\beta\}$, contrary to the hypothesis that $\{1, 2, 4, y, z\}$ is a representative. Thus, $\{y, z\} \subseteq \{6, 7, 11\}$. If $y = 6$ $\{1, 2, 4, 6, z\}\ \alpha = \{1, 2, 3, 4, z\alpha\}$, contrary to hypothesis, so $y = 7$, $z = 11$ must hold, and this completes the proof that there are just four orbits of norm 5 with the representatives listed in Table II.

Finally, consider orbits of norm 3. Since Γ_{12} fixes 1 and 2 and is transitive in 3, 5 and 9, $\{1, 2, x\}$: $x = 3, 5, 9$ are in a single orbit and, similarly, $\{1, 2, y\}$: $y = 4, 6, 7, 8, 10, 11$ are in a single orbit, and this proves that there are (at most) two orbits of norm 3.

The actual listing of the elements of an orbit, *e.g.*, orbit 3.1 in Section IV, is, of course, a slightly more tedious matter. It was made relatively simple by observing that $\{3, 5, 9\}$ is in the same orbit as $\{1, 2, 3\}$, that $\{3, 5, 9\}$ is invariant under Γ_{12}, and that if Λ is a subset of Γ containing 110 members which map the ordered pair (1, 2) on all 110 ordered pairs (i, j), $i \neq j$, then every member of Γ is expressible in the form $\eta\lambda$ where $\eta \in \Gamma_{12}$ and $\lambda \in \Lambda$. Then $\{3, 5, 9\}\ \Gamma = \{3, 5, 9\}\ (\Gamma_{12}\ \Lambda) = (\{3, 5, 9\}\ \Gamma_{12})\ \Lambda = \{3, 5, 9\}\ \Lambda$, so we had only to compute the 110 triples $\{3, 5, 9\}\ \lambda$: $\lambda \in \Lambda$. Λ is readily constructed by choosing a set Λ_1 of 10 members of Γ fixing 1 and mapping 2 on 2, 3, $\cdots$, 10 and a set Λ_2 of 11 members of Γ mapping 1 on 1, $\cdots$, 11. Then it is not difficult to see that $\Lambda = \Lambda_1\Lambda_2$ has the required properties; of course, one computes $\{3, 5, 9\}\ \Lambda_1$ and then applies the members of Λ_2 to these, bypassing the tedious and unnecessary construction of Λ.

Computing ψ's

A *block design*[11] D is an ordered pair of sets A, B with the following properties: A contains a finite number v of objects or "points" a_i, and B consists of b subsets S_i of A called *blocks*. There are positive integral parameters m, r, λ, in addition to v and b, with the following significance:

each S_i has exactly m elements;
each a_i occurs in exactly r blocks; and
each pair $\{a_i, a_j\}$ occurs in exactly λ blocks.

It is not difficult to show that $bm = vr$ and $r(m - 1) = \lambda(v - 1)$. If $b = v$ (and $m = r$), D is said to be a *symmetric* design and in this case we have the following.

Lemma 3:[12] In a symmetric design, two distinct sets have exactly λ objects in common.

Let 0 be an orbit of Γ in S^* and let m be its norm. It follows at once from double transitivity that I, 0 is a block design [with parameters $v = 11$, $b =$ orbit length, $m =$ norm of 0, $r = bm/v$, $\lambda = bm\ (m - 1)/v(v - 1)$]. For example, for orbit 3.1, the parameters are $b = 55$, $m = 3$, $r = 15$, $\lambda = 3$, and we use these facts to compute, *e.g.*, ψ_2 (the sum of orbit 3.1 in the character group) by relatively painless nonarithmetical means as follows. ψ_2 (orbit 1): Take any member of orbit 1, *e.g.*, $\{1\}$. It appears in orbit 3.1 $r = 15$ times so $\psi_2\ (\{1\}) = \psi_2$ (orbit 1) $= -15 + (55 - 15) = +25$.

ψ_2 on orbit $2 = \psi_2\ (\{1, 2\})$ is computed as follows. 1 and 2 each appear $r = 15$ times in orbit 3.1, $\lambda = 3$ of them together. Separate appearance gives -1, simultaneous or nonappearance gives $+1$, so $\psi_2\ (\{1, 2\}) = -24 + 3 + (55\text{–}27) = +7$.

ψ_2 on orbit $3.1 = \psi_2\ (\{1, 2, 3\})$. 1, 2, 3 each appears $r = 15$ times, $3 \cdot \lambda = 9$ times in pairs, once in a triple in orbit 3.1. The triple and solo appearances give -1's, the double and nonappearances give $+1$'s. The triple accounts for 3 doubles, leaving 6 doubles and 30 singles, so $\psi_2\ (\{1, 2, 3\}) = -1 + 6 - 30 + (55\text{–}37) = -7$, and so forth.

Remark: Since the 55 elements of S of type (0, 0, 2) fall into a single orbit (see Table II), we already know that the sum of any two distinct members of rows 13–23 of L has norm 6. The combinatorial origin of this fact becomes clear in the following argument. First observe that orbit 5.4, the complement of the orbit formed by rows 13–23 of L, is a symmetric block design with parameter $\lambda = 2$; hence, by Lemma 3, each pair of its members has two numbers in common. Now let s and t be distinct members of rows 13–23 of L. Then $(s + t)^* = (s^* \cup t^*) \cap \overline{(s^* \cap t^*)} = \overline{(s^* \cap t^*)} \cap (\bar{s}^* \cup \bar{t}^*) =$ the numbers in $\bar{s}^* \cup \bar{t}^*$ which are not common to both; since $\bar{s}^*$ and $\bar{t}^*$ belong to orbit 5.4, each contains three numbers not in the other and so $(s + t)^*$ has 6 members, $||s + t|| = 6$.

The (ψ_1, ψ_3) Decoding Procedure

It is convenient to generalize the notation used for error type as follows: let (i, j, m), where i and m are non-negative integers and $j = 0$ or 1, denote an element of S which is obtained as the sum of i *distinct* members of rows 1–11 of L, j times row 12 and m *distinct* members of rows 13–23 of L. This notation may be used to facilitate arguments which depend on the fact that every nonzero element of S is uniquely expressible as an (i, j, m) with $1 < i + j + m \leq 3$. Thus, *e.g.*, in step 3, if $\varphi(s) = (2, 0, 1)$, we cannot obtain $\varphi(s + e_i) = (1,9)$, for, if u has an error in the ith place, $s + e_i$ is a (1, 0, 1) and $h(s) = (1, 1)$ or $(-3, -3)$, while if u does not have an error in the ith place, $s + e_i$ is a (3, 0, 1); then if $h(s + e_i) = (1, 9)$, $s + e_i$ is also a (0, 1, 1), so we should have, symbolically, (3, 0, 1) = (0, 1, 1), which is equivalent to (3, 0, 0) = (0, 1, 2), or (3, 0, 0) = (0, 1, 0), which are impossible. A slightly less trivial situation is encountered in step 3, *i.e.*, $h(s) = (3, 3)$ or $(-1, -1)$ if $\varphi(s) = (1, 1, 1)$ and for some i, $h(s + e_i) = (1, 1)$. This can occur; the

[11] See Hall [3], p. 59, *et. seq.*
[12] See [3], p. 61, Theorem 1.1.

he (1, 1) indicates not orbit 5.1, which it would if $\varphi(s)$ had been equal to (2, 0, 1), but rather orbit 5.2. We have, symbolically, (2, 1, 1) = (1, 0, 2) which includes (3, 0, 0) = (0, 1, 3), which is certainly possible. Naturally, the decoder could not know at this point that $\varphi(s)$ = (1, 1, 1) rather than (2, 0, 1), although it could avoid the difficulty by searching for $h_i(s)$ = (1, 9) first. However, the decoder performs correctly without the preliminary search, for it would not find j with h_j $(s + e_i)$ = (− 1, −9). Indeed, in order to do so we should have to have (3, 1, 1) = (2, 0, 2) = (0, 0, 1), and the last equation is equivalent to (2, 0, 0) = (0, 0, 3) or (2, 0, 0) = (0, 0, 1), which are impossible. The other facts used in the procedure may be established in a similar way.

BIBLIOGRAPHY

[1] W. Burnside, "Theory of Groups of Finite Order," Dover Publications, Inc., New York, N. Y.; 1955.
[2] M. J. E. Golay, "Notes on digital coding," PROC. IRE, vol. 37, p. 657; June, 1949.
[3] M. Hall, "Projective Planes and Related Topics," California Institute of Technology, Pasadena; April, 1954.
[4] L. J. Paige, "A note on the Mathieu groups," *Can. J. Math.*, vol. 9, pp. 15–18; January, 1956.
[5] E. Prange, "Cyclic Error-Correcting Codes in Two Symbols," AFCRC-TN-57-103, ASTIA Document No. AD 133749; September, 1957.
[6] E. Prange, "Some Cyclic Error-Correcting Codes with Simple Decoding Algorithms," AFCRC-TN-58-156, ASTIA Document No. AD 152386; April, 1958.
[7] D. Slepian, "A class of binary signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956.
[8] E. Prange, "The Use of Coset Equivalence in the Analysis and Decoding of Group Codes," AFCRC-TR-59-164; June, 1959.

Encoding and Error-Correction Procedures for the Bose-Chaudhuri Codes*

W. W. PETERSON†, MEMBER, IRE

Summary—**Bose and Ray-Chaudhuri have recently described a class of binary codes which for arbitrary m and t are t-error correcting and have length $2^m - 1$ of which no more than mt digits are redundancy. This paper describes a simple error-correction procedure for these codes. Their cyclic structure is demonstrated and methods of exploiting it to implement the coding and correction procedure using shift registers are outlined. Closer bounds on the number of redundancy digits are derived.**

INTRODUCTION

BOSE and Chaudhuri[1] have recently discovered a new class of codes with some remarkable properties. For any positive integers m and t, there is a code in this class that consists of blocks of length $2^m - 1$, that corrects t errors, and that requires no more than mt parity check digits. Thus, the codes cover a wide range in rate and error-correcting ability, unlike most other known classes of codes.[2] These codes are a generalization of the Hamming codes;[3] the case $t = 1$ gives the Hamming code in each case.

In this paper two important properties of these codes are described. First, a method for error correction is described which is a generalization of the simple error-correction procedure that can be used with Hamming codes. The procedure requires a number of operations which increases only as a small power of the length of the codes.

Second, it is shown that these are cyclic codes[4] and,

* Received by the PGIT, December 6, 1959. Part of this work was supported by the U. S. Army Signal Corps, the U. S. Air Force Office of Scientific Research, Air Research and Development Command, and the U. S. Navy Office of Naval Research at the Research Laboratory of Electronics, Mass. Inst. Tech., Cambridge, Mass.; and part of the work was done at the IBM Research Lab., Yorktown, N. Y.

† On leave from the University of Florida, Gainesville. Presently at the Dept. of Elec. Engrg. and Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

[1] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error-correcting binary group codes," to be published in *Information and Control.*

[2] The only others of which I am aware are I. S. Reed, "A class of multiple-error-correcting codes and decoding scheme," IRE TRANS. ON INFORMATION THEORY, vol. IT-4, pp. 38–49, September, 1954; P. Elias, "Error free coding," IRE TRANS. ON INFORMATION THEORY, vol. IT-4, pp. 29–37, September, 1954; and I. S. Reed and G. Solomon, "Polynomial code," to be published in *J. Soc. Ind. Appl. Math.*

[3] R. W. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950.

[4] E. Prange, "Some Cyclic Error-Correcting Codes with Simple Decoding Algorithms," Air Force Cambridge Research Center, Bedford, Mass., Tech. Note AFCRC-TN-58-156, April, 1958; "Cyclic Error-Correcting Codes in Two Symbols," Air Force Cambridge Research Center, Bedford, Mass., Tech. Note ARCRC-TN-57-103, September, 1957; "The Use of Coset Equivalence in the Analysis and Decoding of Group Codes," Air Force Cambridge Research Center, Bedford, Mass., Tech. Rept. AFCRC-TR-59-164, June, 1959.

therefore, the encoding can be accomplished very efficiently with a shift register. The theory of the cyclic structure also provides a closer bound on the number of parity checks required to correct a given number of errors.

Construction of the Bose-Chaudhuri Codes

Given an irreducible polynomial $p(X)$ of degree m with 1 and 0 as coefficients, a representation of the Galois Field with 2^m elements $GF(2^m)$ can be formed. It consists of all polynomials of degree $m-1$ or less. They can be added (modulo 2) term by term in the ordinary way. The rule for multiplication is to multiply in the ordinary way, reducing the answer modulo 2 and modulo $p(X)$ to a polynomial of degree $m-1$ or less. (That is, consider $p(X)=0$, and use this equation to eliminate terms of power greater than $m-1$.) It can be shown then that certain of these polynomials, called primitive elements, have the property that the first 2^m-1 powers of such an element are exactly all the 2^m-1 nonzero field elements. Also, every nonzero field element is a root of the equation

$$X^{2^m-1}=1$$

and conversely. Thus if α is any element of the field, $\alpha^{-1}=\alpha^{2^m-2}$.

The field elements can also be thought of as vectors whose components are the coefficients of the polynomials. The sum of two vectors corresponds to the sum of the corresponding polynomials.

The Bose-Chaudhuri codes are described by giving the matrix of parity check rules, which is the matrix

$$M=\begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ \alpha & \alpha^3 & \cdot & \cdot & \cdot & \alpha^{2t-1} \\ \alpha^2 & (\alpha^3)^2 & \cdot & \cdot & \cdot & (\alpha^{2t-1})^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha^{2^m-2} & (\alpha^3)^{2^m-2} & \cdot & \cdot & \cdot & (\alpha^{2t-1})^{2^m-2} \end{bmatrix} \tag{1}$$

where α is a primitive element of the field.

This is a $2^m-1 \times t$ matrix of $GF(2^m)$ elements, but thinking of each field element as a vector of m binary digits, this is a $2^m-1 \times mt$ matrix of binary digits. A vector of 2^m-1 binary digits is considered a code word if it satisfies the parity check described by each column; *i.e.*, if the product of this vector with the matrix is zero. In other words the set of all code words is the (left) null space of this matrix.

The code that Bose and Ray-Chaudhuri use as an example will be used to illustrate the ideas discussed in this paper. Let α denote a root of the equation $X^4=X+1$. This happens to be a primitive element of the field. Then the 15 nonzero field elements are given in Table I.

Taking $t=3$, the following matrix of parity check rules results:

TABLE I
Representation of $GF(2^4)$

α^0	$=1$			$=(1\,0\,0\,0)$
α^1	$=\alpha$			$=(0\,1\,0\,0)$
α^2	$=$	α^2		$=(0\,0\,1\,0)$
α^3	$=$		α^3	$=(0\,0\,0\,1)$
α^4	$=1+\alpha$			$=(1\,1\,0\,0)$
α^5	$=\alpha+\alpha^2$			$=(0\,1\,1\,0)$
α^6	$=$	$\alpha^2+\alpha^3$		$=(0\,0\,1\,1)$
α^7	$=1+\alpha$	$+\alpha^3$		$=(1\,1\,0\,1)$
α^8	$=1$	$+\alpha^2$		$=(1\,0\,1\,0)$
α^9	$=\alpha$	$+\alpha^3$		$=(0\,1\,0\,1)$
α^{10}	$=1+\alpha+\alpha^2$			$=(1\,1\,1\,0)$
α^{11}	$=\alpha+\alpha^2+\alpha^3$			$=(0\,1\,1\,1)$
α^{12}	$=1+\alpha+\alpha^2+\alpha^3$			$=(1\,1\,1\,1)$
α^{13}	$=1$	$+\alpha^2+\alpha^3$		$=(1\,0\,1\,1)$
α^{14}	$=1$	$+\alpha^3$		$=(1\,0\,0\,1)$
α^{15}	$=1=\alpha^0$			

$$M=\begin{bmatrix} 1\,0\,0\,0 & 1\,0\,0\,0 & 1\,0\,0\,0 \\ 0\,1\,0\,0 & 0\,0\,0\,1 & 0\,1\,1\,0 \\ 0\,0\,1\,0 & 0\,0\,1\,1 & 1\,1\,1\,0 \\ 0\,0\,0\,1 & 0\,1\,0\,1 & 1\,0\,0\,0 \\ 1\,1\,0\,0 & 1\,1\,1\,1 & 0\,1\,1\,0 \\ 0\,1\,1\,0 & 1\,0\,0\,0 & 1\,1\,1\,0 \\ 0\,0\,1\,1 & 0\,0\,0\,1 & 1\,0\,0\,0 \\ 1\,1\,0\,1 & 0\,0\,1\,1 & 0\,1\,1\,0 \\ 1\,0\,1\,0 & 0\,1\,0\,1 & 1\,1\,1\,0 \\ 0\,1\,0\,1 & 1\,1\,1\,1 & 1\,0\,0\,0 \\ 1\,1\,1\,0 & 1\,0\,0\,0 & 0\,1\,1\,0 \\ 0\,1\,1\,1 & 0\,0\,0\,1 & 1\,1\,1\,0 \\ 1\,1\,1\,1 & 0\,0\,1\,1 & 1\,0\,0\,0 \\ 1\,0\,1\,1 & 0\,1\,0\,1 & 0\,1\,1\,0 \\ 1\,0\,0\,1 & 1\,1\,1\,1 & 1\,1\,1\,0 \end{bmatrix} \tag{2}$$

Of these twelve columns, the last one is trivial and the next to last is a duplicate; these two can be dropped. The rest are independent, and the result is a code with fifteen digit code words of which ten are parity checks and five are information places. The code corrects all triple errors.

An Error-Correction Procedure

Consider the result of multiplying a vector $(r_0, r_1, r_2, \cdots, r_{n-1})$ of $n=2^m-1$ components by the matrix M in (1). The result is a vector of t Galois field elements. The first component is

$$r_0+r_1\alpha+r_2\alpha^2+\cdots+r_{n-1}\alpha^{n-1}=r(\alpha)$$

where

$$r(X)=r_0+r_1X+\cdots+r_{n-1}X^{n-1}$$

is the polynomial which corresponds naturally to the given vector. (In what follows no distinction will be made

between a vector and the corresponding polynomial.) The other components are clearly $r(\alpha^3), r(\alpha^5), \cdots, r(\alpha^{2t-1})$.

In these terms an equivalent definition of the Bose-Chandhuri codes can be given. A vector is a code word if it is in the left null space of M, *i.e.*, if the parity checks $r(\alpha)$, $r(\alpha^3)$, $r(\alpha^5)$, $\cdots$, $r(\alpha^{2t-1}0)$ are zero. This can be restated as follows:

Definition: A polynomial $s(X)$ is a code vector for a t-error correcting Bose-Chaudhuri code if, and only if, $\alpha, \alpha^3, \cdots, \alpha^{2t-1}$ are roots of $s(X)$.

The first step in devising a decoding method is to characterize the information contained in the parity check calculation for a received vector which may contain errors. Let $e = (e_0, e_1, \cdots, e_{n-1})$ be the vector of errors, *i.e.*, if the errors occur in the positions $i_1, i_2, \cdots, i_v$, then

$$e_i = 1 \text{ for } i = i_1, i_2, \cdots, i_v$$

$$e_i = 0 \text{ otherwise.}$$

There is a one to one correspondence between the elements of the error vector and the elements of $GF(2^m)$ which constitute the first column of the parity check matrix M given by (1), e_i corresponding to the element a^i occurring in the i-th position in the first column of M. The elements $X_1, X_2, \cdots, X_v$ of $GF(2^m)$ which correspond in this way to $e_{i_1}, e_{i_2}, \cdots, e_{i_v}$ may be called the error position numbers. Thus $X_j = a^{i_j}$ $(j = 1, 2, \cdots, v)$.

Lemma 1: If a received vector r has errors in digits numbered $X_1, X_2, \cdots, X_v$, then the parity check vector $r \times M$ is of the form $(S_1, S_3, S_5, \cdots, S_{2t-1})$ where

$$S_j = \sum_{i=1}^{v} X_i^j. \tag{3}$$

Proof: Assume that the vector s was transmitted, and $r = s + e$ received, where e has ones in the positions $i_1, i_2, \cdots, i_v$ and zeros in all other positions. In terms of corresponding polynomials,

$$r(X) = s(X) + e(X)$$

and the result of the parity check calculation is

$$[r(\alpha), r(\alpha^3), \cdots, r(\alpha^{2t-1})].$$

But $s(\alpha) = s(\alpha^3) = \cdots = s(\alpha^{2t-1}) = 0$, so that $r(\alpha) = s(\alpha) + e(\alpha) = e(\alpha)$, $r(\alpha^3) = e(\alpha^3)$, etc. Thus, the result of the parity check calculation is $[e(\alpha), e(\alpha^3), \cdots, e(\alpha^{2t-1})]$. But

$$e(\alpha^j) = e_0 + e_1\alpha^j + e_2\alpha^{2j} + \cdots + e_{n-1}\alpha^{(n-1)j}$$

$$= \sum_{i=1}^{v} \alpha^{i_v j} = \sum_{i=1}^{v} X_i^j \quad \text{Q.E.D.}$$

It is interesting to note that for $t = 1$, if the error occurs, for example, in the component numbered X_1, then the result of the parity check calculation is exactly $S_1 = X_1$ which is the Galois field binary code for the error position number. This is exactly analogous to the method of error-correction for Hamming codes in which the parity check calculation gives the ordinary binary code for the position of the error. In this sense the Bose-Chaudhuri codes for $t = 1$ are equivalent to the Hamming single-error correcting code.

The S_i are the power sum symmetric functions.[5] Thus the parity checks give the first t odd power sum symmetric functions. The first t even ones can be found from the fact that modulo 2, $(a + b)^2 = a^2 + b^2$, and hence

$$S_1^2 = \left[\sum_{i=1}^{v} X_i\right]^2 = \sum_{i=1}^{v} X_i^2 = S_2. \tag{4}$$

Similarly, $S_4 = S_1^4$, $S_6 = S_3^2$, etc.

Suppose that there are t errors. Then the error position numbers $X_1 \cdots X_t$ satisfy the equations

$$S_j = \sum_{i=1}^{t} X_i^j \qquad j = 1, 3, \cdots 2t - 1.$$

This is a set of t equations in t unknowns, the X_i. The solution would tell the positions of the errors. It appears impossible to solve the equations by any direct method, and trying all combinations of t of the $2^m - 1$ field elements would require too many computations. There is, however, an interesting compromise.

The elementary symmetric functions σ_i are related to the power sum symmetric functions S_i by Newton's identities:[5]

$$\left.\begin{aligned} &S_1 - \sigma_1 = 0 \\ &S_2 - S_1\sigma_1 + 2\sigma_2 = 0 \\ &S_3 - S_2\sigma_1 + S_1\sigma_2 - 3\sigma_3 = 0 \\ &S_4 - S_3\sigma_1 + S_2\sigma_2 - S_1\sigma_3 + 4\sigma_4 = 0 \\ &S_5 - S_4\sigma_1 + S_3\sigma_2 - S_2\sigma_3 + S_1\sigma_4 - 5\sigma_5 = 0 \\ &\cdots \text{ etc.} \end{aligned}\right\} \tag{5}$$

If it is possible to solve Newton's identities for the elementary symmetric functions σ_i, the error position numbers must satisfy the equation

$$X^t - \sigma_1 X^{t-1} + \sigma_2 X^{t-2} \cdots \pm \sigma_t$$

$$= (X - X_1)(X - X_2) \cdots (X - X_t) = 0. \tag{6}$$

Eq. (6) can be solved effectively by merely substituting each of the $n = 2^m - 1$ field elements into the equation. For each digit in the received vector, the corresponding $GF(2^m)$ element is substituted in the equation. If the equation is satisfied, this bit is wrong and must be changed. If the equation is not satisfied, the bit is correct.

[5] See, for example, van der Waerden, footnote 8; J. Riordan, "An Introduction to Combinatorial Analysis," John Wiley and Sons, Inc., New York, N. Y., 1958; T. Muir and W. H. Metzler, "A Treatise on the Theory of Determinants," ch. 21, 1930; or any book on the Theory of Equations.

The proof that it is indeed possible to solve for the ordinary symmetric functions from the power sum symmetric functions is given by the following theorem:[6]

Theorem 1: The $k \times k$ matrix

$$M_k = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ S_2 & S_1 & 1 & 0 & \cdots & 0 \\ S_4 & S_3 & S_2 & S_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ S_{2k-4} & S_{2k-5} & S_{2k-6} & S_{2k-7} & & S_{k-3} \\ S_{2k-2} & S_{2k-3} & S_{2k-4} & S_{2k-5} & \cdots & S_{k-1} \end{bmatrix}$$

is nonsingular if power sum symmetric functions S_i are power sums of k or $k - 1$ distinct field elements, and is singular if the S_i are power sums of fewer than $k - 1$ distinct field elements.

The proof requires the following two lemmas:

Lemma 2: If the S_i are power sums of $v \leq k - 2$ distinct field elements, M_k is singular.

Proof:

$$M_k \begin{bmatrix} 0 \\ 1 \\ \sigma_1 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_{k-2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

by Newton's identities, (5), and thus M_k has a nontrivial null space and must be singular. Q.E.D.

Lemma 3: If the S_i are power sums of k indeterminants $X_1, \cdots, X_k$, then the determinant

$$| M_k | = \prod_{j<i} (X_i + X_j).$$

Proof: If $X_i = X_j$, all of the power sums contain two identical terms, which cancel because the field has characteristic 2 (*i.e.*, $2 = 0$). Then it is just as if there were no more than $k - 2$ distinct elements used in forming the power sums, and, by Lemma 2, the determinant is zero. Therefore, $X_i + X_j$ is a factor of the determinant, for all i and j, and the left-hand side must be divisible by the right-hand side. It is easy to check that the left-hand side is homogeneous of degree $k(k - 1)/2$, the same as the right-hand side, and therefore they must differ at most by a constant factor.

To determine the constant factor, a single special case suffices. If k is odd, let the X_i be the roots of the equation

$$X^k - 1 = 0.$$

Then

$$\sum X_i^j = S_j = 0 \quad \text{if} \quad j \equiv 0 \mod t,$$
$$= 1 \quad \text{if} \quad j \equiv 0 \mod t.$$

There will be exactly one 1 in each row and each column and it follows that $|M_k| = 1$ in this case. For k even, letting the X_i be all of the roots of the equation

$$X^k - X = 0$$

gives the same result. The constant factor, which could be only 0 or 1, must be 1.

Now Theorem 1 follows from the fact that if the determinant $|M_k|$ is zero it must be that some $X_i = X_j$. Since all of the nonzero X_i are distinct, $X_i = X_j = 0$, and there were fewer than $k - 1$ errors. Q.E.D.

If there are actually $t - 1$ errors, it can be seen from Newton's identities, Cramer's Rule and Theorem 1 that the solution for the σ's will yield $\sigma_t = 0$. The corresponding polynomial equation will have zero as one root.

Now let us review the error-correcting procedure. The t-error correcting Bose-Chaudhuri codes give, as the parity checks on received sequences, the odd power-sum symmetric functions up to S_{2t-1} and the intermediate even functions can be calculated simply from these. If it is assumed that no more than t errors occur, then by Theorem 1, with $k = t$, it is either possible to solve for the error position numbers, or there are $t - 2$ or fewer errors. In the latter case, $\sigma_{t-1} = \sigma_t = 0$, and two equations can be dropped, giving a set of $t - 2$ equations in $t - 2$ unknowns to which Theorem 1 can be applied again. Eventually, if there were any errors at all, a set of equations that can be solved for the elementary symmetric functions of the error-position numbers will be found.

The correction procedure consists of three phases:

1) calculate the parity checks and the even numbered S_i;

2) from these, calculate the elementary symmetric functions σ_i; and

3) finally, substitute each field element into the equation

$$X^t + \sigma_1 X^{t-1} + \sigma_2 X^{t-2} \cdots + \sigma_t = 0. \tag{7}$$

Those field elements which satisfy this equation correspond to error positions.

The second step involves a certain amount of trial and error because it is possible to solve the equations and obtain correct solutions only when the number of equations used equals or exceeds by one the number of errors that actually occur. This step might be carried out, as an alternative to the procedure described in the preceding paragraph, by starting with the assumption that two errors occurred, solving, and checking the solution. If the solution doesn't check, four errors would be assumed, and so forth. When a set of answers that checks occurs, it must be the correct solution.

[6] Similar results for a real field appear, for example, in H. O. Faulkes, "Theorems of Kakeya and Polya on Power sums," *Math. Z.*, vol. 65, pp. 345–352; 1956.

If it is assumed that the length n of the code approaches infinity and that the number of errors corrected t is a fixed fraction of n, the number of operations required for error correction can be crudely estimated as follows. The first phase, calculating parity checks, requires a number of operations proportional to the number of digits multiplied by the number of parity checks, or no more than nmt operations. This quantity nmt is proportional to n^2 log n. The second phase requires solving a $t \times t$ set of equations. The number of operations for this task is typically proportional to t^3, but it may have to be done $t/2$ times. This will increase in the limit no faster than n^4. Finally, substituting in a t-degree polynomial requires t multiplications and t additions of m digit numbers, and must be done n times, so that $2\ tmn$ is a rough estimate of the number of operations. This again would vary as t^2 log n. Thus, the total number of operations certainly would increase as a small power of n.

Consider, as an example, the code corresponding to the matrix in (2), which corrects triple errors. The appropriate equations are

$$\begin{aligned} &S_1 + \sigma_1 = 0, \\ &S_3 + S_2\sigma_1 + S_1\sigma_2 + \sigma_3 = 0, \quad \text{and} \\ &S_5 + S_4\sigma_1 + S_3\sigma_2 + S_2\sigma_3 = 0. \end{aligned} \tag{8}$$

The parity checks for the received vectors give S_1, S_3, and S_5. $S_2 = S_1^2$, and $S_4 = S_1^4$. Solving for the σ's gives

$$\begin{aligned} &\sigma_1 = S_1, \qquad \sigma_2 = (S_1^2S_3 + S_5)/(S_1^3 + S_3) \quad \text{and} \\ &\sigma_3 = (S_1S_5 + S_3^2 + S_1^3S_3 + S_1^6)/(S_1^3 + S_3), \end{aligned} \tag{9}$$

provided that $S_1^3 + S_3 \neq 0$. If there is only one error $S_1^3 + S_3 = 0$. Furthermore, if $S_1^3 + S_3 = 0$, the Newton's identities yield $\sigma_3 = \sigma_1\sigma_2$, and the equation

$$\begin{aligned} &X^3 + \sigma_1X^2 + \sigma_2X + \sigma_3 \\ &\quad = X^3 + \sigma_1X^2 + \sigma_2X + \sigma_1\sigma_2 \\ &\quad = (X + \sigma_1)(X^2 + \sigma_2) = (X + \sigma_1)(X + \sqrt{\sigma_2})^2 = 0 \end{aligned}$$

has two equal roots, which must be zero, and therefore there is only one error.

As a numerical example, suppose that the vector of all zeros is transmitted, and that errors occur in the 2nd, 5th, and 7th positions. Then

$$r = (0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)$$

$$r \times M = (1\ 0\ 1\ 1 \quad 1\ 1\ 1\ 1 \quad 1\ 0\ 0\ 0)$$

$$S_1 = (1\ 0\ 1\ 1) \qquad S_3 = (1\ 1\ 1\ 1) \qquad S_5 = (1\ 0\ 0\ 0).$$

Referring to Table I, one finds

$$S_2 = S_1^2 = (1\ 0\ 1\ 1)^2 = (\alpha^{13})^2 = \alpha^{26} = \alpha^{11} = (0\ 1\ 1\ 1)$$

and

$$S_4 = S_1^4 = \alpha^{52} = \alpha^7 = (1\ 1\ 0\ 1).$$

Then,

$$\begin{aligned} S_3 + S_1^3 &= (1\ 0\ 1\ 0) \neq 0, \\ \sigma_1 = S_1 &= (1\ 0\ 1\ 1) = \alpha^{13}, \\ \sigma_2 = (S_1^2S_3 + S_5)/S_3 + S_1^3 &= (0\ 0\ 1\ 0)/(1\ 0\ 1\ 0) \\ &= \alpha^2/\alpha^8 = \alpha^9/\alpha^{15} = \alpha^9. \end{aligned}$$

Similarly,

$$\sigma_3 = \alpha^{11}.$$

It is then easy to verify that the equation,

$$X^3 + \alpha^{13}X^2 + \alpha^9X + \alpha^{11} = 0,$$

is satisfied by the three values $X = \alpha$, α^4, and α^6, and only these. These correspond to the errors in r.

Some Properties of Cyclic Codes and Shift Register Generators

Codes for which the code points comprise a cyclic subspace of vectors of zeros and ones have been studied recently by Prange,[4] and, along with theoretical results, he found several efficient codes that can be decoded easily. He has noted that the codes can be coded with the use of a shift-register generator.[7] In this section, some of the theory of cyclic codes and linear recurrent sequences is reviewed briefly from a point of view that is especially well adapted to the study of the Bose-Chaudhuri codes.

A subset C of vectors of n binary digits is called a *cyclic subspace* if it has the following two properties:

1) If v_1 and v_2 are in C, their sum modulo 2 is also in C; that is, C is a subspace, or subgroup; and
2) If $v = (a_0, a_1, \cdots, a_{n-1})$ is in C, the vector $v^1 = (a_{n-1}, a_0, a_1, \cdots, a_{n-2})$ obtained by shifting v cyclically one place is also in C.

Let R_n denote the set of all polynomials

$$a_0 + a_1X \cdots + a_{n-1}X^{n-1}$$

of degree less than n with coefficients 1 and 0. They form a group under modulo 2 addition. Multiplication can be defined modulo $X^n - 1$; that is, these polynomials can be multiplied in the ordinary way, modulo 2, and then reduced again to polynomials of degree less than n by the use of the equation $X^n = 1$. Then R_n is a ring in the mathematical sense. A subset I of R_n is called an *ideal*[8] if it satisfies the following two properties:

1) I is a subgroup of R_n; and
2) if $p(X)$ is in I and $a(X)$ is R_n, then the product $p(X)\ a(X)$ is in I.

---

[7] N. Zierler, "Linear recurring sequences," *J. Soc. Ind. Appl. Math.*, vol. 7, pp. 31–48; March, 1959.

[8] Galois fields and other aspects of algebra used in this paper are treated in many books on modern algebra. See, for example, A. A. Albert, "Fundamental Concepts of Modern Algebra," University of Chicago Press, Chicago, Ill., 1956; G. Birkoff and S. MacLane, "A Survey of Modern Algebra," The Macmillan Co., New York, N. Y., 1953; B. L. van der Waerden, "Modern Algebra," F. Ungar Publishing Co., New York, N. Y., vol. 1 and 2, 1949, 1950.

Considering polynomials $p(X) = a_0 + a_1X \cdots + a_{n-1}X^{n-1}$ to be vectors $(a_0, a_1, \cdots, a_{n-1})$, a cyclic shift is the same as multiplication by X modulo $X^n - 1$. Therefore, *every ideal is a cyclic subspace*. Conversely, if $p(X)$ is in a cyclic subspace C, so is $Xp(X)$. It follows that $X^ip(X)$ must also be in C, and since C is a subspace,

$$\sum_i c_iX^ip(X) = p(X) \sum_i c_iX^i$$

must also be in C. Thus, if $p(X)$ is in C, so is the product of $p(X)$ and any polynomial. Therefore, *every cyclic subspace is an ideal.*

The important but well-known properties of ideals given in the following three lemmas and two theorems are proved here to make the paper self-contained.

Lemma 4: If $p(X)$ and $q(X)$ are in an ideal I, the greatest common divisor (*GCD*), $d(X)$, of $p(X)$ and $q(X)$ is in I.

This follows directly from the fact that it is always possible to express the $d(X)$ in the form

$$d(X) = a(X)p(X) + b(X)q(X)$$

where $a(X)$ and $b(X)$ are polynomials.

Lemma 5: All polynomials in an ideal I are multiples of the unique polynomial of least degree in I. (That is, every ideal is a principal ideal.)

Proof: Let $p(X)$ be a polynomial of least degree in I. Then, if $q(X)$ is any other polynomial in I, the greatest common divisor of $p(X)$ and $q(X)$ is in I. If $p(X)$ does not divide $q(X)$, then the greatest common divisor of $p(X)$ and $q(X)$ would have lower degree than $p(X)$, which is a contradiction. Therefore, every polynomial in I is divisible by $p(X)$. If $p_1(X)$ and $p_2(X)$ both have minimum degree, each must be divisible by the other, and hence they are equal.

The ideal consisting of all multiples of $p(X)$ is denoted $[p(X)]$. The polynomial of least degree in an ideal is called its generator.

Lemma 6: The generator $p(X)$ of an ideal is a factor of $X^n - 1$.

Proof: The *GCD* $d(X)$ of $p(X)$ and $X^n - 1$ can be expressed in the form

$$d(X) = a(X)p(X) + b(X)(X^n - 1)$$

$$\equiv a(X)p(X) \mod X^n - 1;$$

hence, $d(X)$ is in the ideal. But $p(X)$ is divisible by $d(X)$, and since $d(X)$ is in the ideal, $d(X)$ is divisible by $p(X)$. Hence, $p(X) = d(X)$.

These results can be summarized as follows:

Theorem 2: A set of polynomials is an ideal in the ring of polynomials modulo $X^n - 1$ if and only if it consists of all multiples of degree less than n of a factor of $X^n - 1$.

Corollary: If $p(X)$ is a polynomial of degree k which divides into $X^n - 1$, $[p(X)]$ is a vector space of dimension $n - k$.

Proof: The elements of $[p(X)]$ are of the form $c(X)p(X)$ where $c(X)$ is an arbitrary polynomial of degree less than $n - k$. Then the $n - k$ coefficients of $c(X)$ are arbitrary.

Theorem 3: If $p(X)\ q(X) = X^n - 1$, the ideals $[p(X)$ and $[q(X)]$ are null spaces of each other. That is, a polynomial $p_1(X)$ is in $[p(X)]$ if, and only if, $p_1(X)\ q_1(X) =$ modulo $(X^n - 1)$ for every polynomial $q_1(X)$ in $[q(X)$

Proof: Since $p_1(X)$ is in $[p(X)]$, $p_1(X)$ is a multiple o $p(X)$, for example, $a(X)\ p(X)$. Similarly, $q_1(X) = b(X$ $q(X)$. Then $p_1(X)\ q_1(X) = a(X)\ b(X)\ (X^n - 1) =$ Conversely, if $p_1(X)\ q(X) = 0$, then $p_1(X)\ q(X)$ must b a multiple of $X^n - 1$, and $p_1(X)$ must be a multiple o $(X^n - 1)/q(X) = p(X)$.

Note that the fact that the product of two polynomial is zero implies that the dot product of the correspondin two vectors is zero, if in one of them the order of th components is reversed. That is, if

$$(a_0 + a_1X \cdots + a_{n-1}X^{n-1})(b_0 + b_1X \cdots + b_{n-1}X^{n-1}) =$$

then

$$(a_0, a_1 \cdots a_{n-1}) \cdot (b_{n-1}, b_{n-2}, \cdots b_1, b_0)$$

$$= a_0b_{n-1} + a_1b_{n-2} \cdots + a_{n-1}b_0 = 0$$

since this is the coefficient of X^{n-1} in the product of th polynomials. Hence, if $[p(x)]$ and $[q(x)]$ are null space of each other, the corresponding vector-spaces are null spaces of each other provided that the order of component in the vectors of one of these is reversed.

Now let us consider a recursion relation (or differenc equation) of the form

$$\sum_{j=0}^{k} a_jR_{i-j} = 0, \tag{10a}$$

or

$$R_i = \sum_{j=1}^{k} a_jR_{i-j} \qquad a_0 = a_k = 1. \tag{10b}$$

The solution of these equations for given coefficients a will be a sequence of binary digits, $\{R_i\}$. Given the digit $R_0, \cdots, R_{k-1}$, (10) is the rule for calculations R_k, the R_{k+1}, and so forth. Also, the sum of two solutions is agai a solution because the equation is linear. Therefore, th solutions form a vector space of dimension k. The solution are characterized in the following theorem.

Theorem 4: Let $p(X) = \sum_{i=0}^{k} a_iX^i$, $a_0 = a_k = 1$, an let n be the smallest integer for which $X^n - 1$ is divisibl by $p(X)$. Let $q(X) = (X^n - 1)/p(X)$. Then the solution of the difference equation

$$R_i = \sum_{j=1}^{k} a_jR_{i-j}$$

are periodic of period n, and the set made up of the firs period of each possible solution, considered as polynomial is the ideal $[q(X)]$.

Proof: That any vector taken from $[q(X)]$ is a solutio can be seen by multiplying a polynomial from $[q(X)]$, fo example, $q_1(X)$, by $p(X)$. The digits in the product ar formed by the summation in (10a), and, since the produc is zero, (10a) is satisfied. Therefore, any sequence forme by repetition of a vector taken from $[q(X)]$ is a solutio

of (10). Since $q(X) = X^n - 1/p(X)$ has degree $n - k$, then $[q(X)]$ has dimension k, by the corollary to Theorem 2. This is the same as the dimension of the space of solutions, and therefore $[q(X)]$ must include all solutions.

The Cyclic Structure of the Bose-Chaudhuri Codes

It is shown in this section that the Bose-Chaudhuri codes are examples of cyclic codes as studied by Prange.[4] As such they can be generated with very simple equipment, as is illustrated for the (15,5) code in the next section. Out of this theory also comes a better estimate of the number of parity check digits required to correct a given number of errors.

By the alternative definition of the Bose-Chaudhuri codes given in the second section of this paper, a code consists of all polynomials $f(X)$ which have $\alpha, \alpha^3, \cdots, \alpha^{2t-1}$ as roots. Each element α^i of the field is a root of a unique irreducible polynomial $p_i(X)$ of minimum degree. Then $f(X)$ must be divisible by each of the polynomials $p_1(X), p_3(X), \cdots, p_{2t-1}(X)$ and, hence, by their least common multiple:[9]

$$f(X) = \underset{i=1,3,\cdots,2t-1}{\text{LCM}} [p_i(X)]. \tag{11}$$

Since each of the factors $p_i(X)$ is irreducible, the least common multiple of the $p_i(X)$ is simply the product of the polynomials $p_i(X)$, with the duplicates omitted. Duplications are quite possible; they will occur, in fact, for any α^i and α^j that are roots of the same polynomial $p_i(X)$. In other words, should α^i and α^j happen to be roots of the same irreducible polynomial, the columns in the parity check matrix will be dependent, although not necessarily identical. The parity checks produced by the column of powers of α^i will be satisfied if and only if the parity checks produced by the column of powers of α^j are satisfied, and thus one set or the other is unnecessary.

Finally, the set of all sequences that comprise the code can, by Theorem 4, be generated by a recursion relation defined by the polynomial $X^n - 1/f(X)$, and hence by a shift register generator.

At this point it is interesting to study the limiting cases of the minimum and maximum numbers of parity checks. It has already been noted that the nontrivial minimum is the Hamming code. On the other extreme, the last two columns which might be included in the parity check matrix are powers of $\alpha^{2m-2} = \alpha^{-1}$ and $\alpha^{2m-1} = 1$. The last one is a root of the irreducible polynomial $1 + x$ and the resulting code would be the ideal generated by $(1 + x^n)/(1 + x)$. This ideal consists of the zero vector and the vector of all ones, so the code is the trivial repetition of a single information digit $n = 2^m - 1$ times. If α is a primitive element, so is α^{-1}, and therefore the irreducible polynomial of which α^{-1} is a root is primitive. It can be shown then that when only the last two columns, corresponding to α^{-1} and 1, are omitted from the parity check matrix, the resulting code consists of a maximal length sequence, all its shifts, all complements, and a sequence of all 1's, which is then the code studied by San Soucie and Green.[10] This code can also be shown to be equivalent to the Reed-Muller first-order code with any one digit dropped.[11]

It is possible to predict easily which powers of α are roots of the same polynomial, and thus, incidentally, find the degree of the polynomial of which α^i is a root. The method is based on the fact that if a is a root of $f(X)$, then a^2 is also, since $f(a^2) = [f(a)]^2 = 0$. It turns out that $a, a^2, a^4, a^8, \cdots$ are, in fact, all of the roots. In Table II information is given for $m = 4$ and 5. Note that in the first case, $\alpha^{15} = 1$; and in the second, $\alpha^{31} = 1$.

The code for $m = 4$, $t = 3$ has for its generator, by (11),

$$f(X) = p(X)p_3(X)p_5(X)$$

and therefore has $4 + 4 + 2 = 10$ parity checks, and 5 information places. The code for $m = 5$, $t = 5$ has

$$f(X) = p(X)p_3(X)p_5(X)p_7(X)$$

for its generator, and therefore has 20 parity checks. All codes for $m = 4$ and 5 are listed in Table III.

TABLE II
Roots of Polynomials $p_i(X)$

Polynomial		Roots
$m = 4$	$p(X)$	$\alpha, \alpha^2, \alpha^4, \alpha^8$
	$p_3(X)$	$\alpha^3, \alpha^6, \alpha^{12}, \alpha^9$
	$p_5(X)$	α^5, α^{10} $(\alpha^{20} = \alpha^5)$
	$p_7(X)$	$\alpha^7, \alpha^{14}, \alpha^{13}, \alpha^{11}$
$m = 5$	$p(X)$	$\alpha, \alpha^2, \alpha^4, \alpha^8, \alpha^{16}$
	$p_3(X)$	$\alpha^3, \alpha^6, \alpha^{12}, \alpha^{24}, \alpha^{17}$
	$p_5(X)$	$\alpha^5, \alpha^{10}, \alpha^{20}, \alpha^9, \alpha^{18}$
	$p_7(X)$	$\alpha^7, \alpha^{14}, \alpha^{28}, \alpha^{25}, \alpha^{19}$
	$p_9(X) = p_5(X)$	
	$p_{11}(X)$	$\alpha^{11}, \alpha^{22}, \alpha^{13}, \alpha^{26}, \alpha^{21}$
	$p_{13}(X) = p_{11}(X)$	
	$p_{15}(X)$	$\alpha^{15}, \alpha^{30}, \alpha^{29}, \alpha^{27}, \alpha^{23}$

TABLE III
Rate and Error Correction Ability of Bose-Chaudhuri Codes for $m = 4$ and 5

Length of Code Words	Number of Parity Checks	Number of Information Places	Number of Errors Corrected
n	$n - k$	k	t
15	4	11	1
15	8	7	2
15	10	5	3
31	5	26	1
31	10	21	2
31	15	16	3
31	20	11	5
31	25	6	7

[9] See, for example, Birkhoff and MacLane, *op. cit.*, p. 396.

[10] J. H. Green, Jr. and R. L. San Soucie, "An error-correcting encoder and decoder of high efficiency," Proc. IRE, vol. 46, pp. 1741–1744; October, 1958.

[11] N. Zierler, "On a variation of the first-order Reed-Muller Codes," Lincoln Laboratory Group Rept. 34–80; October, 1958.

Code parameters for some larger codes were calculated on the IBM 704 computer. The results are plotted in Fig. 1. The vertical axis represents rate (percentage of all digits available for information), and the horizontal axis represents the number of errors correctable as a percentage of the total number of digits. The dashed curve represents asymptotic values of a lower bound on the rate of the best code that corrects errors in a given percentage of the digits.[12] The curves drawn for the Bose-Chaudhuri codes for large n fall below the bound for the best code. In fact, it is shown in the Appendix that they approach zero as the length of the code increases indefinitely. This may mean that these codes are truly not optimum, or it may mean that the number of errors correctable by the procedure given in this paper is not the total number of errors correctable by Bose-Chaudhuri codes in the case of very long codes.[13]

The polynomial $p(X)$ can be any primitive polynomial of degree m. The other polynomials $p_i(X)$ are determined by the particular choice of $p(X)$, and the question arise as to how they may be calculated. One simple method i based on the fact that every element of $GF(2^m)$ is a roo of the polynomial $X^{2^m-1} - 1$. Therefore, each element i a root of one of the factors of $X^{2^m-1} - 1$. One needs only to factor this polynomial and test to see which factor has X^i as a root. The following alternative method is useful It has been noted that the degree m_i of $p_i(X)$ can be easily determined. Then if

$$p_i(X) = a_0 + a_1X + \cdots + a_{m-1}X^{m_i-1} + X^{m_i},$$

since α^i is a root of $p_i(X)$,

$$0 = a_0\alpha^0 + a_1\alpha^1 + \cdots + a_{m-1}\alpha^{m_i-1} + \alpha^{m_i},$$

and if α^i is written as a vector with m components, the resulting set of linear equations can be solved for the coefficients a_i of $p_i(X)$.

There is also an explicit formula

$$p(X^{1/j})p(\alpha X^{1/j}) \cdots p(\alpha^{j-1}X^{1/j})$$

where α is a primitive jth root of unity. It can be shown that when the multiplication is carried out only integral powers of X remain, and these have only ones or zeros as coefficients.

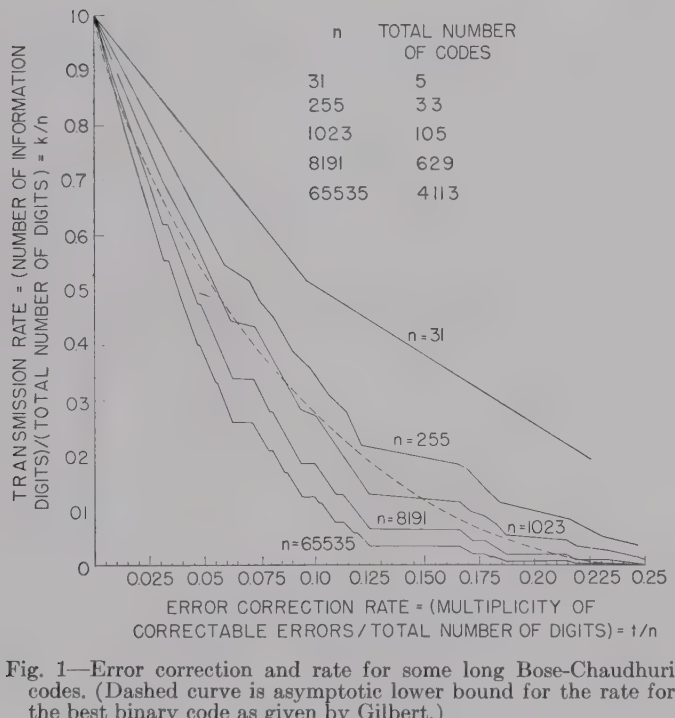


Fig. 1—Error correction and rate for some long Bose-Chaudhuri codes. (Dashed curve is asymptotic lower bound for the rate for the best binary code as given by Gilbert.)

Consider again the sample code discussed by Bose and Chaudhuri. The irreducible factors of $X^{15} - 1$ are

$$X^{15} - 1 = (X - 1)(X^2 + X + 1)(X^4 + X^3 + X^2 + X + 1)(X^4 + X^3 + 1)(X^4 + X + 1)$$

A root of the last factor was taken as α; and thus

$$p(X) = X^4 + X + 1.$$

Then α^3 satisfies the equation $X^5 - 1 = 0$, since $\alpha^{15} = 1$. But $X^5 - 1 = (X - 1)(X^4 + X^3 + X^2 + X + 1)$, and since α^3 is not a root of the first factor, it must be a root of the second. Similarly, α^5 satisfies $X^3 - 1 = 0 = (X - 1)(X^2 + X + 1)$, and so α^5 is a root of $X^2 + X + 1$. The fact that this has degree 2 ties in with the observation that the column of powers of α^5 contained only two independent parity checks.

All code points must be multiples, then, of

$$\begin{aligned} f(X) &= p(X)p_3(X)p_5(X) \\ &= (1 + X + X^4)(1 + X + X^2 + X^3 + X^4) \cdot (1 + X + X^2) \\ &= 1 + X + X^2 + X^4 + X^5 + X^8 + X^{10} \\ &= (1\ 1\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0), \end{aligned} \tag{12}$$

and it can easily be checked that this vector, any cyclic permutation of it, and any sum of permutations, actually do satisfy the parity checks defined by the matrix M in (2).

[12] E. N. Gilbert, "A comparison of signaling alphabets," *Bell Sys. Tech. J.*, vol. 31, pp. 504–522; May, 1952.

[13] I have found with the aid of the IBM 704 that the Bose-Chaudhuri two-error correcting codes for $m = 4$ and 5 correct some triple errors and nothing beyond and are therefore optimum. The three-error correcting code for $m = 4$ corrects 420 quadruple and 28 quintuple error patterns and is optimum. The three-error correcting code for $m = 8$ corrects 13,020 quadruples and 14,756 quintuples and nothing beyond—this seems good but has not been proved optimum. (See A. B. Fontaine and W. W. Peterson, "Group code equivalence and optimum codes," IRE Trans. on Information Theory, vol. IT-5, pp. 60–70; May, 1959.) Thus, any non-optimum behavior of these codes occurs only in codes so large that they are difficult to analyze by looking at code words themselves or searching for coset leaders even with the aid of a computer.

Mechanizing the Coding and Error-Correction

Shift registers with feedback corrections can be used in a number of ways in mechanizing coding and error-correction procedures. The following uses will be discussed in this section:

1) coding using a shift register with one stage for each information digit in the code,
2) coding using a shift register with one stage for each parity check digit in the code,
3) counting in the Galois field code,
4) multiplying and dividing Galois field elements, and
5) calculating parity checks on received vectors.

Both the methods of coding apply to any cyclic code. The methods will be illustrated using the Bose-Chaudhuri (15, 5) code described by the matrix M in (2).

Every cyclic code is an ideal generated by some polynomial $f(X)$, *i.e.*, a polynomial is a code vector if and only if it is divisible by $f(X)$. This means that, by Theorem 4, a vector is a code vector if and only if it satisfies the recursion relation corresponding to the polynomial $(X^n - 1)/f(X)$. For the code used as an example, by (12),

$$f(X) = 1 + X + X^2 + X^4 + X^5 + X^8 + X^{10}$$

$$(1 - X^{15})/f(X) = 1 + X + X^3 + X^5.$$

Then every sequence satisfying the recursion relation

$$R_i = R_{i-1} + R_{i-3} + R_{i-5}$$

is a code point, and conversely. Such sequences can be generated by putting information digits in the shift register generator shown in Fig. 2 and shifting 15 times. The first five digits coming out will be information digits, and the next ten digits will be a set of parity checks which make the whole sequence a code point. The symbols come out of this encoder low order digits first. The order can be reversed by reversing the order of the shift register feedback connections.

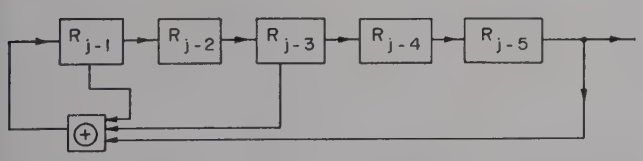


Fig. 2—A shift register for encoding the Bose-Chaudhuri (15,5) code.

A second method of coding is based again on the fact that the coded vector must be, considered as a polynomial, a multiple of $f(X)$. Let $t_0(X)$ be a polynomial in which the k coefficients of the terms involving X^{n-1} through X^{n-k} are arbitrary information digits, and the coefficients of lower order terms are zero. This corresponds to a vector in which the first $n - k$ components are zero, the last k digits arbitrary information digits. Then $t_0(X)$ can be divided by $f(X)$ to produce a quotient and a remainder

$$t_0(X) = f(X)q(X) + r(X),$$

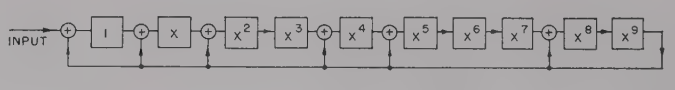
where $r(X)$ has degree less than $(n - k)$, which is the degree of $f(X)$. Then

$$t_0(X) + r(X) = f(X)q(X)$$

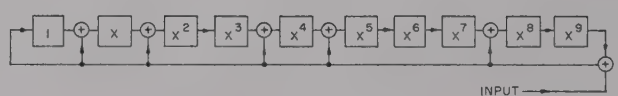
and, hence, $t_0(X) + r(X)$ is a code point. But $r(X)$ corresponds to a vector in which all components except the first $n - k$ are zero, since $r(X)$ has degree less than $n - k$. Thus, the sum consists of $n - k$ check digits, the coefficients of $r(X)$, and k information digits, the coefficients of $t_0(X)$.

The next problem is to calculate $r(X)$. In general, the calculation of the remainder after division by a polynomial can be accomplished with a shift register. The method is illustrated in Fig. 3(a). Assuming the divisor is the $f(X)$ for the code used in the example, *i.e.*, $1 + X + X^2 + X^4 + X^5 + X^8 + X^{10}$, the operation of the circuit can be understood as follows: The answer is the same as results from reducing the dividend modulo $f(X)$. This means that the dividend polynomial should be reduced to a polynomial of degree less than 10 using the relation

$$X^{10} = 1 + X + X^2 + X^4 + X^5 + X^8. \qquad (13)$$



(a)



(b)

Fig. 3—Shift register for calculating residues modulo $f(X) = 1 + X + X^2 + X^4 + X^5 + X^8 + X^{10}$. (a) Basic circuit; (b) basic circuit with automatic premultiplication by X^{10}.

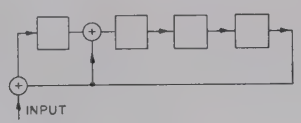
Now assume that a single one is shifted into the low-order position and then shifted right a number of times. Thinking of the contents of the register as a polynomial with low order digits at the left, each shift corresponds to multiplying by X, at least until a shift out of the high-order position. A one in the high-order position corresponds to X^9, and shifting it out makes it X^{10}. This results in the circuit in adding into the lower order positions the equivalent of X^{10} given in (13), and, hence, in this case the shift still corresponds to multiplying by X and modulo $f(X)$. Thus, successive shifts give successive powers of X modulo $f(X)$.

Now this is a linear device, and a polynomial (which is the sum of powers of X) can be reduced modulo $f(X)$ by shifting it into the device, high power terms first, until the constant term is shifted into the low-order position.

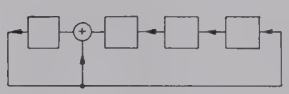
In using this device for calculating the $r(X)$ in (17), the modification shown in Fig. 3(b) can be made to avoid the last $n - k$ shifts which would add $n - k$ zeros into the low-order positions. It amounts to multiplying the input digits by $X^{n-k} = X^{10}$ before adding.

The procedure for coding is then to shift all the information digits into the device in Fig. 3(a) or 3(b). If the device in Fig. 3(a) is used, $n - k$ more shifts must be made with no input. Then the correct check digits remain in the register and should simply follow the information digits, high order digits first, to make a complete code vector. Note that the number of stages in this shift register is $n - k$, while the shift register shown in Fig. 2 has k stages.

A counter which counts in terms of Galois field elements is shown in Fig. 4(a). It works on the same principle as the device shown in Fig. 3(a), but using the primitive polynomial $p(X) = X^4 + X + 1$ of which α is a root. If a 1 is placed in the low-order position, successive shifts give successive powers of α using the relation $\alpha^4 = \alpha + 1$, and these are exactly the representations of $GF(2^4)$ elements given in Table I.



(a)



(b)

Fig. 4—Galois field counters for $GF(2^4)$. (a) Increasing powers of α; and (b) decreasing powers of α.

In the device shown in Fig. 3(b), a left shift corresponds to division by α and a 1 shifted out of the low order end α^{-1} is replaced by its equivalent $1 + \alpha^3$. Thus, this device can count down, or give Galois field elements in reverse order. A multiplier can be mechanized by putting one factor in a device A like that shown in Fig. 3(a), the other in a device B like that shown in Fig. 3(b). Then both devices are shifted until the code for 1 appears in device B. The product then appears in A. Division can be done in an anologous manner. Multiplication can also be done in a manner analogous to that used in digital computers with a shift register such as that shown in Fig. 3(a) used in place of an accumulator.

The parity checks corresponding to the first column of Galois field elements in the matrix M of (2) correspond to the Galois field representation of

$$r(\alpha) = r_0 + r_1\alpha + r_2\alpha^2 \cdots r_{2^m-2}\alpha^{2^m-2}.$$

This can be calculated by using the relation $\alpha^4 + \alpha + 1 = 0$ to eliminate terms of degree higher than 3 in α. This, in turn, is exactly what will result if the vector $(r_0, r_1, \cdots, r_2m_2)$ is shifted into the shift register shown in Fig. 3(a) high-order digits first. Note that shifting fifteen times multiplies by α^{15}, but $\alpha^{15} = 1$. Similarly, the device in Fig. 3(b) could be used with the low-order digits entering first.

Calculation of the other parity checks is slightly more complicated. It requires calculating $r(\alpha^j)$ for the first t odd values of j. The first step is to devise a shift register which automatically multiplies by α^j. The example $j = 5$ should make the principles clear. Note that

$$1 \cdot \alpha^5 = \alpha^5 = \alpha + \alpha^2$$
$$\alpha \cdot \alpha^5 = \alpha^6 = \alpha^2 + \alpha^3$$
$$\alpha^2 \cdot \alpha^5 = \alpha^7 = 1 + \alpha + \alpha^3$$
$$\alpha^3 \cdot \alpha^5 = \alpha^8 = 1 + \alpha^2,$$

so that

$$\begin{aligned}\alpha^5(a_0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3) &= a_0(\alpha + \alpha^2) + a_1(\alpha^2 + \alpha^3) \\ &\quad + a_2(1 + \alpha + \alpha^3) + a_3(1 + \alpha^2) \\ &= (a_2 + a_3) + (a_0 + a_2)\alpha \\ &\quad + (a_0 + a_1 + a_3)\alpha^2 + (a_1 + a_2)\alpha^3.\end{aligned}$$

Thus, the new value of a_0 is the old $a_2 + a_3$, the new a_1 is the old $a_0 + a_2$, etc. A shift register with feedback connections shown in Fig. 5 will give this result. Then, if the received vector $(r_0, r_1, \cdots, r_2m_2)$ is shifted into this device, after fifteen shifts the result $r(\alpha^5)$ will remain in the register.

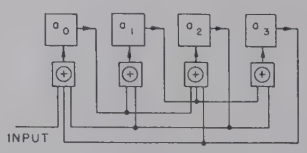


Fig. 5—A circuit for calculating the parity checks $r(\alpha^5)$.

Conclusion

Relatively simple coding and error-correcting methods have been described for the Bose-Chaudhuri codes. The study of coding and error-correction methods for these codes gives additional insight into the remarkable structure of the codes.

Appendix

A bound on the rate of Bose-Chaudhuri codes which correct $t = 2^\lambda$ errors is derived in this Appendix, and it is shown on the basis of this bound that if t is made a fixed fraction of n, the number of digits in the code, the rate must approach zero as n increases indefinitely.

This problem is purely number-theoretic, and can be formulated as follows: The quantity to be studied is the rate, which is the quotient of the number k of information digits and $n = 2^m - 1$, the total number of digits. Since there is one independent parity check for each distinct residue of $j2^i$ for $1 \leq j \leq 2t$, $0 \leq i < m$, the number of such residues in $n - k$. Since $2^m = 2^0$ modulo $2^m - 1$, the condition $0 \leq i < m$ can be replaced by $1 \leq i \leq m$. For convenience in what follows, j will be allowed to take on the value zero also; this adds one distinct residue.

Let $N(s)$ be the number of distinct residues of $j2^i$ for $0 \leq j < 2t = 2^{\lambda+1}$ and $m - s < i \leq m$. Then

$$n - k = N(m) - 1 \geq N(s) - 1 \quad \text{if} \quad s \leq m$$

and

$$k = n - N(m) + 1$$

$$= 2^m - N(m) \leq 2^m - N(s) \quad \text{if} \quad s \leq m. \tag{14}$$

An equation for $N(s)$, valid only for $s \leq \lambda$, will be derived but this will give an upper bound on k by (14).

Consider first the residues for a particular value of i, $m - \lambda \leq i \leq m$. They can be arranged as follows:

$$\begin{array}{rrrcr}
0\cdot 2^i, & 1\cdot 2^i, & 2\cdot 2^i, & \cdots, & (2^{m-i}-1)2^i \\
(2^{m-i}+0)2^i, & (2^{m-i}+1)2^i, & (2^{m-i}+2)2^i, & \cdots, & (2\cdot 2^{m-i}-1)2^i \\
(2\cdot 2^{m-i}+0)2^i, & (2\cdot 2^{m-i}+1)2^i, & (2\cdot 2^{m-i}+2)2^i, & \cdots, & (3\cdot 2^{m-i}-1)2^i \\
\vdots & \vdots & \vdots & & \vdots \\
(2^{\lambda+1}-2^{m-i}+0)2^i, & (2^{\lambda+1}-2^{m-i}+1)2^i, & (2^{\lambda+1}-2^{m-i}+2)2^i, & \cdots, & (2^{\lambda+1}-1)2^i.
\end{array}$$

In this array there are $2^{\lambda+1-m+i}$ rows. Since $2^m \equiv 1$, the array can be rewritten

$$\begin{array}{rrrcr}
0, & 1\cdot 2^i, & 2\cdot 2^i, & \cdots, & (2^{m-i}-1)2^i \\
1, & 1+1\cdot 2^i, & 1+2\cdot 2^i, & \cdots, & 1+(2^{m-i}-1)2^i \\
2, & 2+1\cdot 2^i, & 2+2\cdot 2^i, & \cdots, & 2+(2^{m-i}-1)2^i \\
\vdots & \vdots & \vdots & & \vdots \\
2^{\lambda+1+i-m}-1, & (2^{\lambda+1+i-m}-1)+1\cdot 2^i, & (2^{\lambda+1+i-m}-1)+2\cdot 2^i, & \cdots, & 2^{\lambda+1+i-m}-1+(2^{m-i}-1)2^i.
\end{array}$$

This consists exactly of 2^{m-i} sets of $2^{\lambda+1+i-m}$ successive numbers starting at each multiple of 2^i. The arrangement is shown graphically in Fig. 6.

The important facts can be seen clearly in Fig. 6 but are tedious to prove formally. For each i there are $2^{\lambda+1}$ residues and therefore, in particular, $N(1) = 2^{\lambda+1}$. Two adjacent columns in Fig. 6 have half their residues in common. In particular, $N(2) = 2^{\lambda+1} + 2^\lambda$. Now in adding the contributions to $N(s)$ for larger values of s it is necessary to determine exactly how many residues have occurred in all previous columns combined. There is one other case which must be considered besides the previous adjacent column. Note that the residues and nonresidues of $j\cdot 2^i$ for a particular value of i fall in blocks of $2^{\lambda+1+i-m}$ successive numbers. In determining which residues for a particular value of i, for example, i_0, have occurred before, each block of 2^{i_0+1} successive numbers is treated the same. Each will have two blocks of $2^{\lambda+1+i_0-m}$ residues. The first will already have been counted in the $i_0 + 1^{st}$ column. The fraction of the others to be omitted is the same as the fraction of blocks of length 2^{i_0+1} which were counted as residues for $i \geq i_0 + m - \lambda$, which is the same as $N(\lambda - i_0)/2^m$. Then, since $s = m - i_0$,

$$N(s) = N(s-1) + 2^\lambda\cdot[1 - 2^{-m}N(\lambda - m + s)] \tag{15}$$

for $0 < s \leq \lambda$. [$N(s)$ should be considered zero for $s \leq 0$.]

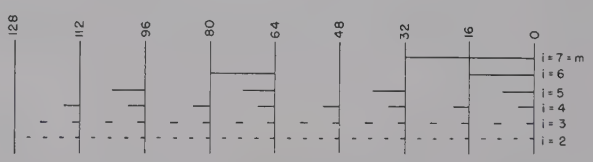


Fig. 6—Distribution of residues of $j2^i (m = 7, \lambda = 4)$.

Now let

$$R(s) = 1 - N(s)\cdot 2^{-m}.$$

Since $N(s)$ includes the zero residue, the actual number of parity digits is at least $N(s) - 1$. The actual number of information digits is at most $2^m - 1 - N(s) + 1 = 2^m - N(s)$. The actual rate would be at most $[2^m - N(s)]/(2^m - 1)$, but for large m, this is approximately $R(s)$. Then

$$N(s) = 2^m[1 - R(s)],$$

and substitution in (15) results in a difference equation for $R(s)$:

$$R(s) = R(s - 1) - 2^{\lambda-m}R(s - m + \lambda) \qquad (16)$$

for $0 < s$. [$R(s)$ should be considered to be 1 for $s \leq 0$.] Clearly,

$$1 \geq R(s) \geq 0 \quad \text{for all} \quad s. \qquad (17)$$

It follows at once from (16) and (17) that $R(s)$ is nonincreasing. Now if there exists $\epsilon > 0$ such that $R(s) > \epsilon$ for all s, choose any $s_0 > m - \lambda + (2^{m-\lambda}/\epsilon)$. Then $R(s_0) = [R(s_0) - R(s_0 - 1)] + [R(s_0 - 1) - R(s_0 - 2)] + \cdots + [R(m - \lambda + 1) - R(m - \lambda)] + R(m - \lambda)$ trivially $= R(m - \lambda) - 2^{\lambda-m}[R(s_0 - m + \lambda) + R(s_0 - m + \lambda -) + \cdots + R(1)]$ by (16) $< R(m - \lambda) - 2^{\lambda-m}(s_0 - m + \lambda)\epsilon$ by hyoothesis $< R(m - \lambda) - 1$ by choice of $s_0 < 0$ by half of (17), contradicting the other half, and proving that $R(S) \to 0$ as $s \to \infty$ must hold.

Now suppose that it is required that errors be corrected in a fraction 2^{-v} of the number of digits in a code word. Then

$$2^{-v} = 2^{\lambda}/2^m - 1 \approx 2^{\lambda-m},$$

so $v \approx m - \lambda$. Then, taking $s = \lambda$, $R(\lambda) = R(m - v)$ is an upper bound on the rate for a code with $2^m - 1$ digits. As m increases this approaches zero. Since rate is a monotone nonincreasing function of the number of errors correctable and the rate approaches zero for arbitrarily small fractions $t/n = 2^{-v}$, it must approach zero for any fraction $t/n > 0$.

Acknowledgment

I have benefited greatly from discussions with many people at the IBM Research Laboratory and at the Research Laboratory of Electronics at Massachusetts Institute of Technology. E. Prange, of the Air Force Cambridge Research Center, J. Griesmer and J. Selfridge of IBM, and M. P. Schutzenberger and S. Golomb at the Research Laboratory of Electronics were especially helpful.

Most of all, I am indebted to R. C. Bose of the University of North Carolina, for lecturing on his and Chaudhuri's fine work so soon after it was done and for the very stimulating discussion we had during his visit to the IBM Research Laboratory in August 1959.

Part of the computation work was done at the M.I.T. Computation Center.

Synchronization of Binary Messages*

E. N. GILBERT†

Summary—When messages are transmitted as blocks of binary digits, means of locating the beginnings of blocks are provided to keep the receiver in synchronism with the transmitter. Ordinarily, one uses a special synchronizing symbol (which is really a third kind of digit, neither 0 nor 1) for this purpose. The Morse code letter space and the teletype start and stop pulses are examples. If a special synchronizing digit is not available, its function may be served by a short sequence of binary digits P which is placed as a prefix to each block. The other digits must then be constrained to keep the sequence P from appearing within a block. If blocks of N digits (including the prefix P) are used, the prefix should be chosen to make large the number $G(N)$ of different blocks which satisfy the constraints. Lengthening the prefix decreases the number of "message digits" which remain in the block but also relaxes the constraints. Thus, for each N, there corresponds some optimum length of prefix.

For each prefix P, a generating function, a recurrence formula, and an asymptotic formula for large N are found for $G(N)$. Tables of $G(N)$ are given for all prefixes of four digits or fewer. Among all prefixes P of a given length A, the one for which $G(N)$ has the most rapid growth is $P = 11 \cdots 1$. However, for this choice of P, the table of values of $G(N)$ starts with small values; $11 \cdots 1$ does not become the best A-digit prefix until N is very large. At these values of N, the $(A + 1)$ − digit prefix $11 \cdots 10$ is still better. The tables suggest that, for any N, a best prefix can always be found in the form $11 \cdots 10$, for suitable A. Taking $P = 11 \cdots 10$ and $A = [\log_2(N \log_2 e)]$ it is shown that $G(N)$ is roughly $0.35N^{-1}2^N$. This result is near optimal since no choice of P can make $G(N)$ exceed $N^{-1}2^N$.

* Received by the PGIT, December 20, 1959.
† Mathematical Res. Dept., Bell Telephone Labs., Murray Hill, N. J.

I. Introduction

WHEN block coding is used, some care must be taken to ensure that the transmitter and receiver stay in synchronism. For example, the Morse code letter spaces and the teletype stop and start pulses

are used to mark the beginnings of new letters. Without some synchronizing scheme, a receiver turned on in the middle of a message might start decoding in the middle of a letter and emit gibberish.

In binary systems it is rarely practical to use one of the digits as a synchronizer. If 1 were used for this purpose, the only available codes[1] would be 1, 10, 100, 1000, $\cdots$. As an alternative, one might rely on the self-synchronizing ability of a suitable variable-length binary encoding [3]. These encodings bring the receiver into synchronism after some delay and have the advantage that no time is wasted sending synchronizing information. However, the delay to achieve synchronism depends on the message being transmitted and so is somewhat unpredictable; occasional long delays may be encountered.

Redundancy may be used to provide an encoding in which the synchronization delay is always held below a fixed limit. The comma-free encodings of Golomb, Welch, and Delbrück [5] and Golomb, Gordon, and Welch [4] are of this kind. These encodings have codes of fixed length, say N digits. The encoding is a list of N-tuples (codes) so chosen that, if the receiver starts to decode when it is out of synchronism, then it always sees an N-tuple which is not one of the codes in the list. After at most $N - 1$ digit times, and at most $N - 1$ false starts, the receiver finds the correct synchronism.

A simple example of a comma-free encoding with $N = 5$ is the list of six 5-tuples:

0 1 0 0 0

0 1 1 0 0

0 1 0 1 0

0 1 1 1 0

0 1 0 1 1

0 1 1 1 1.

This encoding has redundancy $1 - (\log_2 6)/5 = 0.48$. It will be shown that very small redundancies are possible when N is large. All the encodings to be considered are comma-free encodings of a special kind called *prefix synchronized encodings*. A particular A-tuple ($A < N$) is selected and called a *synchronizing prefix*. Each code has the synchronizing prefix as its first A digits. The remaining $N - A$ digits of the codes are chosen so that, in an encoded message, no block of A consecutive digits can agree with the synchronizing prefix except blocks of A digits taken at the beginnings of codes. For example, if the synchronizing prefix is taken to be 1010, then 10101101100 is an allowed code but not 10100101011, 10101001101, nor 10101101110. Then the synchronizing prefix is closely analogous to the sync pulse of the teletype encoding. Since only A digits, instead of N, need be remembered in determining synchronism, prefix synchronized encodings may be slightly easier to mechanize than comma-free encodings in general.

The number $G(N)$ of different N-tuples which such encodings can have depends on the choice of the prefix. Recurrence formulas, tables, and asymptotic formulas for $G(N)$ are contained herein. For fixed A, and sufficiently large N, the prefix which maximizes $G(N)$ is $11\cdots1$. However, if A may be varied, $11\cdots1$ is never as good as a suitably chosen longer prefix. Tables support the conjecture that $G(N)$ is always maximized by a prefix of the form $11\cdots10$. All comma-free encodings have redundancy at least as great as $(\log_2 N)/N$. It is shown that this bound is approached, for large N, by suitable prefix synchronized encodings.

II. State Diagrams

Suppose that a synchronizing prefix P, consisting of A binary digits $p_1, p_2, \cdots, p_A$ has been chosen. Each code is constructed by choosing n more binary digits $x_1, x_2, \cdots, x_n$ to make an N-tuple ($N = A + n$)

$$(p_1, p_2, \cdots, p_A, x_1, x_2, \cdots, x_n).$$

The allowed choices of $x_1, \cdots, x_n$ are those for which no A consecutive digits taken from the $(N + A - 2)$-tuple

$$(p_2, \cdots, p_A, x_1, x_2, \cdots, x_n, p_1, \cdots, p_{A-1})$$

agree with the A-tuple P.

These constraints may be reinterpreted graphically. Imagine a conceptual machine which will scan an incoming message digit by digit and ring a bell whenever the synchronizing prefix P appears. Of course, such a machine is easily designed using a shift register to remember the A most recent digits. However some of the 2^A states of the shift register may be merged together. A machine with only $A + 1$ states $S_0, S_1, \cdots, S_A$ may be described as follows.

Let M denote the A-tuple formed by the A most recent message digits. If $M = P$, the machine is to be in state S_0 and the bell must ring.

For $k \geq 1$, the machine is to be in state S_k if both of the following conditions 1) and 2) hold:

1) The last $A - k$ digits of M are the first $A - k$ digits of P.
2) For no integer k' in $0 \leq k' < k$ does 1) hold with k replaced by k'.

Thus, in state S_k, at least k more digits must arrive before the bell can ring. For example, if $P = 0101$ and $M = 1100$, then the state is S_3; the three digits 101 must follow M to produce P. More generally with $P = 0101$, the correspondence between M and the state of the machine is given in Table I. In Table I, x's represent digits which may be either 0 or 1; for example, S_2 corresponds to 0001, 1001, and 1101.

State diagrams of machines for three choices of P are drawn in Fig. 1. The states $S_0, S_1, S_2, \cdots$ are represented

[1] The word *code* will be used for any one of certain strings of binary digits which are allowed to be transmitted. The collection of all codes is called an *encoding*. This usage permits a distinction which is not commonly made (as in "The code for E is a dot, in Morse code").

TABLE I

State	M
S_0	$0101 = P$
S_1	$x010$
S_2	$xx01$ but not 0101
S_3	either $xx00$ or $x110$
S_4	$xx11$

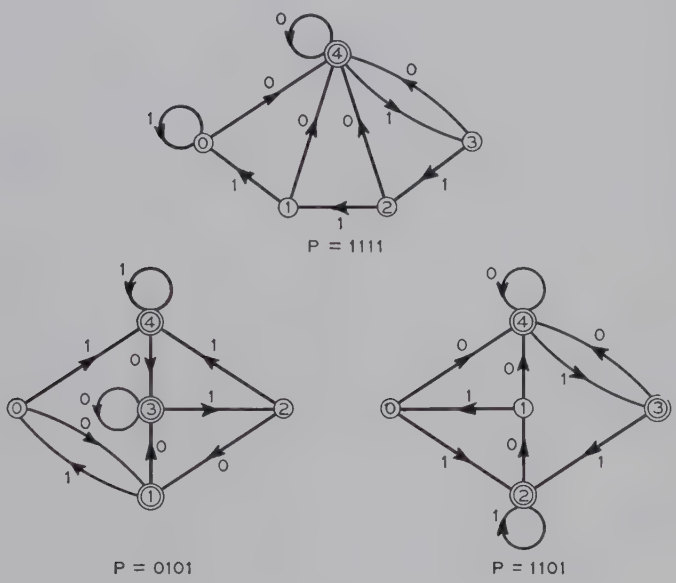


Fig. 1.

by nodes labeled 0, 1, 2, $\cdots$. The transitions which may occur when a new digit is received are shown as arrows. The labels 0, 1 on the arrows denote the value of the new digit which is required to cause the transition. It is not difficult to verify that the states S_0, S_1, $\cdots$ so defined do describe a valid machine (see Appendix I for details).

Given a state S, a path which begins at S and follows arrows of the state diagram may be associated with the sequence of binary digits which is encountered on the arrows. For example, in Fig. 1 with $P = 0101$, the path which starts at 3 and visits the states 2, 4, 4, 3, 3, 2, 1, 0, 4, in that order, is associated with the binary sequence 111001011. This association provides a graphical way of stating the constraints which have been placed on the digits $x_1, \cdots, x_n$ of an allowed code. Starting from the state S_0 the path $x_1, x_2, \cdots, x_n, p_1, \cdots, p_{A-1}$ must never return to S_0. Such a return would indicate an appearance of the A-tuple P in the block of digits

$$p_2, p_3, \cdots, p_A, x_1, \cdots, x_n, p_1, \cdots, p_{A-1}.$$

Alternatively, let a state S ($S \neq S_0$) be called an *end state* if the path $p_1, p_2, \cdots, p_{A-1}$, starting at S, never visits S_0. Then the path $x_1, \cdots, x_n$, starting at S_0, must never return to S_0 and must end at an end state. In Fig. 1 the end states appear as double circles.

III. Numbers of Codes

The list of all possible N-tuples (codes) of the form $(p_1, \cdots, p_A, x_1, \cdots, x_{N-A})$ in which the x's satisfy the requirements of Section II is a comma-free encoding. In this section the number $G(N)$ of such codes will be found.

$G(N)$ depends not only on N but also on the A-tuple prefix P. Fortunately, all 2^A possible prefixes do not have different functions $G(N)$. Two simple symmetry transformations may be applied to a prefix P to obtain new prefixes which have the same $G(N)$ as P. One symmetry is *complementation*, which replaces each digit p_i of P by $1 - p_i$. The second symmetry is *reversal*, which rewrites the digits of P in reverse order (p_i is replaced by p_{A+1-i}). Applied to $P = 01101$, complementation and reversal produce 10010 and 10110. To see that these transformations leave $G(N)$ unchanged, one may note that if digits $x_1, \cdots, x_n$ are allowed to follow a prefix P, then $1 - x_1, \cdots, 1 - x_n$ may follow the complement of P and $x_n, \cdots, x_1$ may follow the reversal of P. Two different prefixes P, P' will be called *equivalent* if P' can be obtained by applying a complementation, or a reversal, or both to P.

The 2^A prefixes may now be collected into a smaller number of classes of equivalent prefixes. Since equivalent prefixes have the same $G(N)$, it suffices to compute $G(N)$ for one prefix from each equivalent class. Using Polya's theorem [6], the number of equivalence classes of A-tuples turns out to be

$$\tfrac{1}{4}\{2^A + 2^{[(A+2)/2]}\}$$

where the straight brackets [ ] denote "integer part." For $A = 2, 3, 4$ every A-tuple is equivalent to one of the following: 11, 10, 111, 110, 101, 1111, 0111, 1101, 0110, 0011, 0101.

Some numerical results appear in Table II. The best prefix P [*i.e.*, the one with the largest $G(N)$] has a curious dependence on N. If A is fixed at $A = 3$, then the choice $P = 110$ is best until $N = 14$, $P = 101$ is best for $N = 15, \cdots, 19$, and $P = 111$ is best thereafter. For any fixed A, the prefix $11\cdots1$ is ultimately best (this will be proved in Theorem 2). However, Table II shows, even for $A = 4$, that other prefixes are better until N becomes very large.

If a best prefix is sought without fixing A, then a prefix $11\cdots10$ with suitable A is always obtained from Table II. It is conjectured that a similar result holds for any value of N. To test this conjecture, J. B. Kruskal extended the table to $N \leq 134$ and $A = 6$ on the IBM 704 computer. These computations support the conjecture. The best choice of A and the corresponding maximum $G(N)$ are given in Table III. This table also lists the number $[N^{-1}\, 2^N]$ which, for all P, is an upper bound on $G(N)$ (see Section V). For $N \leq 28$ the tabulated values of $G(N)$ exceed 2^{N-6}. Since prefixes with $A \geq 6$ have $G(N) \leq 2^{N-6}$, the values $G(N)$ given for $N \leq 28$ are proved best. Comparison with the extended table proves the remaining seven values best. The $G(N)$ values for $N > 35$ were computed using eight place accuracy. When $N \geq 43$ the six-digit prefix 111110 is better than 11110 (the corresponding values of $G(43)$ are 7.9709×10^{10} and 7.9627×10^{10}).

To compute $G(N)$, for a particular choice of an A-tuple prefix P, an Ath order recurrence was used. This re-

TABLE II
$G(N)$ FOR DIFFERENT PREFIXES

	11	10	111	110	101	1111	1110 and 1100	1010	1101 and 1001
$N = 3$	1	2							
4	1	3	1	2	1				
5	2	4	1	4	2	1	2	2	2
6	3	5	2	7	4	1	4	3	3
7	5	6	4	12	7	2	8	4	6
8	8	7	7	20	12	4	15	9	11
9	13	8	13	33	21	8	28	18	21
10	21	9	24	54	37	15	52	32	39
11	34	10	44	88	65	29	96	60	73
12	55	11	81	143	114	56	177	115	136
13	89	12	149	232	200	108	326	216	254
14	144	13	274	376	351	208	600	405	474
15	233	14	504	609	616	401	1104	764	885
16	377	15	927	986	1081	773	2031	1440	1652
17	610	16	1705	1596	1897	1490	3736	2710	3084
18			3136	2583	3329	2872	6872	5103	5757
19			5768	4180	5842	5536	12640	9612	10747
20			10609	6764	10252	10671	23249	18101	20062
21						20569	42762	34086	37451
22						39648	78652	64192	69912

TABLE III
$G(N)$ FOR $P = 11 \cdots 10$

N	A	$G(N)$	$[N^{-1}2^N]$	N	A	$G(N)$	$[N^{-1}2^N]\times 10^{-6}$
3	2	2	2	19	4	12,640	0.027,6
4	2	3	4	20	4	23,249	0.052,4
5	2 or 3	4	6	21	4	42,762	0.100
6	3	7	10	22	5	82,392	0.190
7	3	12	18	23	5	158,816	0.365
8	3	20	32	24	5	306,128	0.70
9	3	33	56	25	5	590,081	1.34
10	3	54	102	26	5	1,137,418	2.58
11	4	96	186	27	5	2,192,444	4.96
12	4	177	341	28	5	4,226,072	9.6
13	4	326	630	29	5	8,146,016	18.5
14	4	600	1170	30	5	15,701,951	35.8
15	4	1,104	2180	31	5	30,266,484	69.5
16	4	2,031	4096	32	5	58,340,524	134
17	4	3,736	7710	33	5	112,454,976	258
18	4	6,872	14563	34	5	216,763,936	505
				35	5	417,825,921	981

currence will be derived in Theorem 1 [see (2)]. In preparation for the theorem some notation will be explained here.

Let $F(n) = G(A + n)$. The theorem gives a formula [see (1)] for a generating function

$$f(z) = \sum_{n=0}^{\infty} F(n)z^n.$$

Indeed, $G(N)$ might also have been computed, with somewhat more difficulty, by expanding $f(z)$ [as given by (1)] in a power series to get the coefficient of z^{N-A}. Theorem 1 will also mention some parameters $V_1, \cdots, V_A, T_1, \cdots, T_{A-1}$. These numbers count certain kinds of paths in the state diagram for the prefix P.

Let T_n be the number of binary n-tuples which describe paths in the state diagram starting at S_0 and ending at S_0. Let V_n denote the number of binary n-tuples describing paths from S_0 to one of the end states. T_n and V_n both count paths which may revisit S_0. The convention $T_0 = 1$, $V_0 = 0$ will be adopted. The numbers T_n and V_n are easily found by direct enumeration. For example, taking $P = 1101$,

$$T_0 = 1, \quad T_1 = 0, \quad T_2 = 0, \quad T_3 = 1, \quad T_4 = 1,$$
$$V_0 = 0, \quad V_1 = 2, \quad V_2 = 3, \quad V_3 = 6, \quad V_4 = 13.$$

In general, when $n \geq A$, T_n counts all n-tuples for which the last A digits form the prefix P; then,

$$T_n = 2^{n-A}, \qquad n \geq A.$$

Likewise,

$$V_n = V_A 2^{n-A}, \qquad n \geq A.$$

When $T_1, \cdots, T_{A-1}, V_1, \cdots, V_A$ have been found, $G(N)$ may be computed with the aid of the following theorem.

Theorem 1: *For $N \geq A$, $G(N)$ is the coefficient of z^{N-A} in the power series of the generating function*

$$f(z) = \frac{(1 - 2z)(V_1 z + \cdots + V_{A-1}z^{A-1}) + V_A z^A}{(1 - 2z)(1 + T_1 z + \cdots + T_{A-1}z^{A-1}) + z^A}. \tag{1}$$

For $N > 2A$, $G(N)$ satisfies a recurrence

$$G(N) + d_1 G(N - 1) + \cdots + d_A G(N - A) = 0 \tag{2}$$

where

$$d_k = \begin{cases} T_k - 2T_{k-1}, & k = 1, 2, \cdots, A - 1 \\ 1 - 2T_{A-1}, & k = A. \end{cases}$$

Proof: $T_k F(n - k)$ of the V_n paths from S_0 to an end state visit S_0 for the last time at the kth step. It follows that

$$\sum_{k=0}^{n} T_k F(n - k) = V_n.$$

Introducing generating functions $t(z) = \sum T_n z^n$ and $v(z) = \sum V_n z^n$, this relationship provides $f(z) = v(z)/t(z)$. Simple formulas may be written for $v(z)$ and $t(z)$. For example,

$$v(z) = V_0 + V_1 z + \cdots + V_{A-1}z^{A-1} + \frac{V_A z^A}{1 - 2z}.$$

Then, $f(z)$ is expressed as a rational function

$$f(z) = \frac{(1 - 2z)(V_0 + V_1 z + \cdots + V_{A-1}z^{A-1}) + V_A z^A}{(1 - 2z)(T_0 + T_1 z + \cdots + T_{A-1}z^{A-1}) + z^A} \tag{3}$$

which proves (1).

Let the numerator and denominator polynomials of (3) be called $E(z)$ and $D(z)$. The denominator polynomial, when multiplied out, is $D(z) = 1 + d_1 z + \cdots + d_A z^A$. By (3) the function $D(z)f(z)$ is a polynomial $E(z)$ of degree $\leq A$. Setting the coefficient of z^n in $D(z)f(z)$ equal to zero one finds the recurrence

$$F(n) + d_1 F(n - 1) + \cdots + d_A F(n - A) = 0 \tag{4}$$

for $n > A$, which proves (2).

For example, taking $P = 1101$, the parameters V_i and T_i were computed earlier. Then the theorem states that

$$f(z) = \frac{(1 - 2z)(2z + 3z^2 + 6z^3) + 13z^4}{(1 - 2z)(1 + z^3) + z^4}$$

$$= \frac{2z - z^2 + z^4}{1 - 2z + z^3 - z^4}$$

and

$$G(N) = 2G(N - 1) - G(N - 3) + G(N - 4), \quad N > 8.$$

Certain pairs of prefixes have the same set of values of $V_1, \cdots, V^A, T_1, \cdots, T_{A-1}$. Then, by (1), they have the same $G(N)$. This explains why the pairs 1110, 1100 and 1101, 1001 are tabulated only once in Table II.

If exact values beyond the range of Table II are needed they may be obtained by the recurrence (4). The coefficients in (4) appear in the denominators of the functions $f(z)$ listed in Table IV.

TABLE IV
GENERATING FUNCTIONS OF $F(n)$

Prefix	$f(z)$
11	$z/(1 - z - z^2)$
10	$(2z - z^2)/(1 - 2z + z^2)$
111	$z/(1 - z - z^2 - z^3)$
110	$(2z - z^3)/(1 - 2z + z^3)$
101	$(z + z^3)/(1 - 2z + z^2 - z^3)$
1111	$z/(1 - z - z^2 - z^3 - z^4)$
1110 and 1100	$(2z - z^4)/(1 - 2z + z^4)$
1010	$(2z - z^2)/(1 - 2z + z^2 - 2z^3 + z^4)$
1101 and 1001	$(2z - z^2 + z^4)/(1 - 2z + z^3 - z^4)$

IV. Asymptotic Formulas

When n is large $F(n)$ grows exponentially. The rate of growth is determined by the poles of $f(z)$ which have the smallest absolute value. If all poles of $f(z)$ lie outside a circle $|z| \leq 1/w$ then $F(n)$ is of order $o(w^n)$. Since the coefficients $F(n)$ of $f(z)$ are positive, one of the poles of $f(z)$ of smallest absolute value lies on the positive real axis (see [7], Chapter VII). Typically, one finds that the smallest positive real pole, for example, at $z = 1/W$, is a simple pole and that there are no other poles of absolute value $1/W$. In that case, it follows from (3) that

$$F(n) \sim \{-E(1/W)/D'(1/W)\}W^{n+1} \tag{5}$$

asymptotically for large n. For, under the conditions stated, the function

$$h(z) = f(z) + E(1/W)/\{(z - 1/W)\ D'(1/W)\}$$

has only the poles of $f(z)$ other than $1/W$. Since these poles have magnitudes greater than $1/W$, the nth coefficient of the series for $h(z)$ is $o(W^n)$ and then (5) follows.

Since poles of $f(z)$ are zeros of the denominator polynomial $D(z)$, W is the largest real positive root of $D(1/W) = 0$. Equivalently, $t(1/W) = 0$ where

$$t(z) = 1 + T_1 z + \cdots + T_{A-1}z^{A-1} + \frac{z^A}{1 - 2z}. \tag{6}$$

Since $F(n) \leq 2^n$, only the range $W \leq 2$, or $z \geq \frac{1}{2}$ need be considered. At $z = \frac{1}{2}$, $t(z)$ has a simple pole with $t(\frac{1}{2} + 0) = -\infty$. For larger z, the term $z^A/(1 - 2z)$ and hence also $t(z)$, increases monotonically. At the value $z = A/(2A - 2)$,

$$t(z) \geq 1 + \frac{z^A}{1 - 2z} = 1 - \frac{A - 1}{\left(2 - \frac{2}{A}\right)^A}.$$

If $A \geq 3$, then $2 - 2/A \geq 4/3$ so that

$$t\left(\frac{A}{2A - 2}\right) \geq 1 - (A - 1)(3/4)^A > 0.$$

Then, for $A \geq 3$, the real pole is a simple one and

$$2\left(1 - \frac{1}{A}\right) < W < 2. \tag{7}$$

In one case, when $A = 2$ ($P = 10$), the pole occurs at 1 and is a double pole; then (5) does not apply.

It is of some interest to find those prefixes P of length A which produce the largest and the smallest values of W.

Theorem 2: Of all A-digit prefixes P, the one which makes G(N) grow most rapidly for large N is P = 11···1. Slowest growth of G(N) for large N is obtained with any one of the A − 1 prefixes 11···10, 11···100, ···, 10···0. However, for all N > A + 1, the number of codes G(N) obtained with the A-digit prefix 11···1, is never as great as the number obtained with the (A + 1) − digit prefix 11···10.

Proof: In the range defined by (7), $t(z)$ is a monotone increasing function of both z and $T_1, \cdots, T_{A-1}$. Then, an increase in one of the T_k will require a decrease in the value of z which satisfies $t(z) = 0$. It follows that the largest W is obtained if $T_1 = \cdots = T_{A-1} = 1$ and the smallest W is obtained if $T_1 = \cdots = T_{A-1} = 0$. The former case corresponds to the prefix $P = 1 \cdots 1$. The latter corresponds to any of the prefixes $111\cdots10$, $11\cdots100, \cdots, 10\cdots0$.

Although when A is fixed, the choice $P = 111\cdots1$ produces the most rapid growth of $F(n)$ for large n, this choice is never the best one when N is fixed and A can be varied. For example, every N-tuple which is allowed for the prefix $11\cdots1$ (A ones) is also allowed for the prefix $11\cdots10$ (A ones and 1 zero). The converse is not true; for example, when $A = 3$ and $N = 10$, the 10-tuple

1110101111

is allowed for the prefix 1110 but not for 111. Then $11\cdots10$ is always a better prefix than $11\cdots1$.

As an application, let $D_A(z)$ denote the denominator polynomial for the prefix $11\cdots1$ (A ones):

$$D_A(z) = 1 - z - z^2 - \cdots - z^A.$$

Let $1/W_A$ be the smallest real positive root of $D_A(z) = 0$. The denominator polynomial for the prefix $11\cdots10$ ($A - 1$ ones and 1 zero) is

$$1 - 2z + z^A = (1 - z)\ D_{A-1}(z);$$

then for the prefix $11\cdots10$, $W = W_{A-1}$. Theorem 2 now shows for all prefixes of length A, that

$$W_{A-1} \leq W \leq W_A, \tag{8}$$

which is an improvement over (7).

When $A \geq 2$ all zeros of $D_A(z)$, except $z = 1/W_A$, lie outside the unit circle; see Appendix II for a proof. Then, taking $P = 11\cdots1$, not only does (5) apply but the error $\to 0$ as $n \to \infty$. The integer $F(n)$ becomes just the closest integer to the expression (5) when n is large. Of course, to use (5) in this manner W_A must be computed to high accuracy; the recurrence (3) is still the better way of computing $F(n)$ exactly.

Similarly, taking $P = 11\cdots10$, the two poles $z = 1/W_{A-1}$ and $z = 1$ of $f(z)$ supply an asymptotic formula ($A \geq 3$)

$$F(n) \sim \frac{W_{A-1}^{n+1}}{2 - AW_{A-1}^{1-A}} + \frac{1}{2 - A}.$$

The dominant term in this formula is given by (5) with $W = W_{A-1}$, $E(z) = 2z - z^A$, $D(z) = 1 - 2z + z^A$; note that $E(1/W_{A-1}) = 1$ since $1/W_{A-1}$ is a zero of the polynomial $1 - 2z + z^A = 1 - E(z)$. The constant term $1/(2 - A)$ is contributed by the pole at $z = 1$ and is also obtained from (5), now with $W = 1$. Again, the nearest integer to the number given by the formula is the exact value of $F(n)$ when n is large. The following theorem gives a bound which holds for all $N > A$.

Theorem 3: *If the synchronizing prefix is the A-tuple* $11\cdots10$, *then*

$$G(N) > (W_{A-1})^{N-A}, \qquad N > A. \tag{9}$$

Proof: Write the generating function in the form

$$f(z) = \frac{2z - z^A}{1 - 2z + z^A}$$

$$= -1 + \frac{1}{(1 - z)\,D_{A-1}(z)}$$

or

$$D_{A-1}(z)f(z) = \frac{1}{1 - z} - D_{A-1}(z)$$

$$= 2z + \cdots + 2z^{A-1} + z^A + z^{A+1} + \cdots.$$

Equating coefficients of z^n, one finds

$$F(n) = 1 + F(n - 1) + \cdots + F(n - A + 1) \tag{10}$$

for $n > A$. Now the bound (9) will follow by induction on n. When $n = 1, \cdots, A - 1$,

$$F(n) = 2^n > (W_{A-1})^n.$$

If $n \geq A$ and if (9) holds for $1, 2, \cdots, n - 1$,

$$F(n) > 1 + (W_{A-1})^{n-1} + \cdots + (W_{A-1})^{n-A+1}$$

by (10). But, since $D_{A-1}(1/W_{A-1}) = 0$,

$$F(n) > 1 + (W_{A-1})^n,$$

which verifies (9).

In applying (8) and (9) a series for W_A is useful.

Theorem 4: *If* $A \geq 2$, *then*

$$W_A = 2\left\{1 - \sum_{n=1}^{\infty} \frac{2^{-nA-n}}{n}\binom{nA + n - 2}{n - 1}\right\}. \tag{11}$$

A proof is given in Appendix III.

V. Redundancy Estimates

Delbrück, Golomb, Gordon and Welch [4, 5] obtain an upper bound on the number of codes which a comma-free encoding may contain. This bound is a number of equivalence classes of periodic sequences with least period N. Two periodic sequences are considered equivalent if they differ only in phase (for example $\cdots110110110\cdots$ and $\cdots101101101\cdots$ are equivalent). An exact formula for this number of classes is given in [4] and [5]; see also [2]. For our purposes, a simpler bound

$$G(N) \leq 2^N/N \tag{12}$$

suffices. This bound follows from the bound of Delbrück, Golomb, Gordon, and Welch when it is noted that each sequence of least period N is equivalent to just $N - 1$ other sequences and that no more than 2^N sequences have least period N. In Table III the best values of $G(N)$ fall short of $2^N/N$ by factors of the order of $1/2$.

From (12) it is clear that a redundancy at least as great as $(\log_2 N)/N$ is necessary for a comma-free encoding using N-tuples.

Redundancies of the same small order of magnitude may be achieved using the prefix synchronized encoding, as will be shown presently. First, however, a very simple example will be given in which the redundancy is roughly $2N^{-1/2}$. The synchronizing prefix will be $111\cdots1$ (A digits) and $N = A^2 + 1$. Instead of using all N-tuples which satisfy the constraints of Section II, the stronger constraints

$$x_1 = x_{A+1} = x_{2A+1} = \cdots = x_{N-A} = 0$$

are imposed. Thus, the typical N-tuple has the appearance (for $A = 4$)

$$11110xxx0xxx0xxx0$$

where the x's represent digits which are unrestricted. Since $2A$ of the $A^2 + 1$ digits are fixed, the redundancy is $2A/(A^2 + 1)$, which equals $2(N - 1)^{\frac{1}{2}}/N$. A possible advantage of this encoding is that the unrestricted digits (x's) may be taken directly from a given binary message without further encoding.

The following theorem cites a family of prefix synchronized encodings which have redundancies approximating $(\log_2 N)/N$.

Theorem 5: *Choose the synchronizing prefix to be* $P = 11\cdots10$ *with*

$$A = [\log_2 (N \log_2 e)].$$

Then a constant C exists such that

$$G(N) > \frac{2^N}{N}\left(\frac{1}{2\log_2 e}\right)\left(1 - \frac{C\log_2 N}{N}\right).$$

Proof: The bound will follow from (9) using a suitable estimate of W_{A-1}. It follows from (11), (see Appendix III) that

$$W_{A-1} = 2\{1 - 2^{-A} + 0(A2^{-2A})\}$$
$$= 2e^{-2^{-A}+0(N^{-2}\log_2 N)}.$$

Then the lower bound (9) becomes

$$G_N > 2^{N-A}e^{-2^{-A}(N-A)}\{1 + 0(N^{-1}\log_2 N)\}. \quad (13)$$

Let a real number a be defined by

$$A = [\log_2 (N\log_2 e)] = \log_2 (N/a). \quad (14)$$

Then (13) becomes

$$G_N > N^{-1}2^N a e^{-a}\{1 + 0(N^{-1}\log_2 N)\}. \quad (15)$$

By (14), it follows that

$$\frac{1}{\log_2 e} \leq a < \frac{2}{\log_2 e} \quad (16)$$

and this result may now be restated in the form

$$ae^{-a} \geq (2\log_2 e)^{-1}. \quad (17)$$

To derive (17) from (16), note that the function $a\,e^{-a}$ grows monotonically from 0 at $a = 0$ to e^{-1} at $a = 1$, and then decreases monotonically for $a > 1$. At both end-points of the interval (16), $a\,e^{-a}$ has the value $(2\log_2 e)^{-1}$; then (17) is satisfied within the interval. Now the factor $a\,e^{-a}$ in (15) may be bounded by (17) to prove the theorem.

For large N, Theorem 5, shows that the upper bound $N^{-1}2^N$ on $G(N)$ can be achieved to within a constant factor $(2\log_2 e)^{-1} = 0.346$.

Appendix I

State Diagrams

The states $S_0, S_1, \cdots, S_A$, defined in Section II have been obtained by merging some of the 2^A states of the shift register machine. It remains to verify that these combined states are indeed states of a valid machine. Consider two input sequences, one ending in an A-tuple M, the other ending in an A-tuple M', both of which put the machine in state S_k. It must be shown that if both M and M' are followed by the same next digit d, then the two sequences ending in Md and $M'd$ correspond to the same new state.

Suppose $2 \leq k \leq A - 1$. To correspond to S_k, the last $A - k$ digits of both M and M' must be $p_1, p_2, \cdots, p_{A-k}$. If $d = p_{A-k+1}$, then S_{k-1} is the new state following both M and M'. If d is not p_{A-k-1} then the new states for Md and $M'd$ are S_k and $S_{K'}$ where both $k \leq K$ and $k \leq K'$. However the sequence of K digits which will lead from Md to the A-tuple P will also lead from $M'd$ to P; thus $K' \leq K$. Similarly $K \leq K'$. Then $K = K'$, which was to be proved. The cases $k = 1$ and $k = A$ may be handled by a similar argument.

Appendix II

Zeros of $D_A(z)$

The only zero of $D_A(z)$ in the unit circle $|z| \leq 1$ is the real zero at $z = 1/W_A$. This result is obtained by applying Rouché's theorem [1] to the polynomial $(1 - z)D_A(z) = 1 - 2z + z^{A+1}$. At a point $z = e^{iu}$ on the unit circle,

$$\left|\frac{z^{A+1}}{1-2z}\right| = \frac{1}{|1 - 2e^{iu}|} = \frac{1}{5 - 4\cos u}$$

except at the angle $u = 0 (z = 1)$. In the neighborhood of $z = 1$, say at $z = 1 - b$,

$$\left|\frac{z^{A+1}}{1-2z}\right| = |1 - (A-1)b + 0(b^2)|.$$

It follows that $|z^{A+1}/(1 - 2z)| < 1$ everywhere on a contour C consisting of the unit circle with a small indentation to the left of the point $z = 1$. Then, inside C, $1 - 2z + z^{A+1}$ and $1 - 2z$ have the same number of zeros, namely only one.

Appendix III

Formula for W_A

W_A is the real root of $W = 2 - W^{-A}$ in the interval (7). The substitution $W = 2(1 - x)$ leads to

$$x(1 - x)^A = 2^{-A-1}.$$

This equation will be solved as a special case $w = 2^{-A-1}$ of the equation

$$x(1 - x)^A = w.$$

Using Lagrange's inversion formula [1], a power series about the point $w = 0$ is found for x:

$$x = \sum_{n=1}^{\infty}\frac{w^n}{n!}\left[\frac{d^{n-1}}{dz^{n-1}}(1-z)^{-An}\right]_{z=0}$$
$$= \sum_{n=1}^{\infty}\frac{w^n}{n}(-1)^{n-1}\binom{-An}{n-1}.$$

Formula (11) is obtained by setting $w = 2^{-A-1}$. The substitution is allowable if the series for x converges at this value of w. To check convergence, examine the ratio R_n of the $(n+1)^{st}$ term to the nth in the series for x:

$$R_n = w\frac{n(A+1)-1}{n+1}\prod_{k=0}^{A-1}\frac{n(A+1)+k}{nA+k}$$
$$= w\frac{A+1-1/n}{1+1/n}\prod_{k=0}^{A-1}\left\{\frac{A+1}{A} - \frac{k}{A(nA+k)}\right\},$$
$$R_n < R_\infty = wA^{-A}(A+1)^{A+1}.$$

When $w = 2^{-A-1}$ and $A \geq 2$,

$$R_\infty \leq \frac{A+1}{2}\left(\frac{2}{3}\right)^A < 1$$

and the series converges.

In the proof of Theorem 5, an estimate is needed for the error made in approximating W_A by the first few terms of the series (11). Since $R_n < R_\infty$, the error committed is no more than $(1 - R_\infty)^{-1}$ times the first neglected term. For large A, the factor $(1 - R_\infty)^{-1}$ approaches 1.

References

[1] E. T. Copson, "An Introduction to the Theory of Functions of a Complex Variable," Oxford University Press, New York, N. Y.; 1934.
[2] N. J. Fine, "Classes of periodic sequences," *Illinois J. Math.*, vol. 2, pp. 285–302; June, 1958.
[3] E. N. Gilbert and E. F. Moore, "Variable-length binary encodings," *Bell Syst. Tech. J.*, vol. 38, pp. 933–967; July, 1959.
[4] S. W. Golomb, B. Gordon, and L. R. Welch, "Comma-free codes," *Canadian J. Math.*, vol. 10, no. 2, pp. 202–209; 1958.
[5] S. W. Golomb, L. R. Welch, and M. Delbrück, "Construction and properties of comma-free codes," *Biol. Med. Danske Vid. Selsk.*, vol. 23, no. 9, pp. 1–34; 1958.
[6] John Riordan, "An Introduction to Combinatorial Analysis," John Wiley and Sons, Inc., New York, N. Y.; 1958.
[7] E. C. Titchmarsh, "The Theory of Functions," 2nd ed., Oxford University Press, New York, N. Y.; 1939.

Analytic Inversion of a Class of Covariance Matrices*

WILLIAM A. JANOS†, MEMBER, IRE

Summary—**The sample covariance matrix arising out of finite memory linear least squares estimation over a set of equally spaced time points, is inverted by spectral methods (operationally referred to as the z transform). It is shown that the complexity of the problem depends only upon the complexity of the input correlation function. The final solution is shown to reduce to the inversion of a triangular system of linear equations of an order less than half the degree of the denominator of the input power spectral density function.**

List of Basic Symbols

These are the different symbols used to denote functions, variables, constants, and possibly unconventional mathematical operations used in the presentation. The same symbol with appropriate sub or superscripts may be used for different purposes. Indices of summation, dummy variables and common mathematical symbols have been omitted.

- C = with subscripts, coefficient of exponential in autocorrelation function;
- D = number of exponentials in autocorrelation function;
- M = filter memory, number of intervals T long;
- N = degree of factor of numerator of spectral density functions;
- P = polynomial of degree not greater than $D - 1$, with prime also;
- Q = with subscripts, primed and bar, polynomial of degree $D - N - 1$;
- q = with sub and superscripts, coefficients of Q (above);
- R = with bar and prime, polynomial of degree less than N;
- r = with sub and superscripts, coefficients of R (above);
- T = sampling interval;
- v = defined to be zero over memory interval $(0, MT)$;
- V^* = z transform of v;
- W = optimal weighting sequence, star superscript denotes its z transform;
- ϕ = autocorrelation function;
- Φ^* = transform of ϕ, spectral density function;
- φ = factor of numerator or denominator of Φ^* depending on N or D subscripts;
- α = with subscripts, decay factor in exponentials of ϕ, without subscript used in the Example at the end of this paper to condense notation;
- β = signal autocorrelation function decay factor used in the Example;
- γ = noise autocorrelation function decay factor used in the Example;
- z = spectral term, in "z" transform;
- χ = with subscripts, coefficients of $D - 1$ degree polynomial in z^{-1};
- Ψ = arbitrary right side of Wiener Hopf equation;
- ψ = with subscripts, coefficient of $D - 1$ degree polynomial in z^{-1}; and
- σ = with subscripts, root mean signal or noise power.

* Received by the PGIT, December 6, 1959.
† Raytheon Co., Wayland, Mass.

The following symbols are defined in context and are used to condense notation.

A
B
E
F
G
H
k
p
f

Introduction

IN cases of finite memory least squares estimation of some linear functional of a discrete, uniformly sampled stationary time series, the optimalizing condition is expressed by the integral equation [1–3] (in the Stieltjes sense);

$$\sum_{n=0}^{M} W(nT)\phi[(m-n)T] = \psi(mT), \qquad m \in (0, M) \tag{1}$$

$W(nT)$ is the weighting sequence or filter to be determined; $\phi(nT)$, the autocorrelation function of the input signal plus noise, is assumed known and consisting of a linear combination of exponentials [4];

$$\phi(mT) = \sum_{k=1}^{D} C_k e^{-\alpha_k T|m|}, \qquad Re\,\alpha_k > 0, \tag{2}$$

and $\psi(mT)$ may be assumed arbitrary.

The solution of (1) may be obtained by considering a related problem. To determine $W_\mu(nT)$ such that

$$\sum_{0}^{M} W_\mu(nT)\phi[(m-n)T] = \delta_{\mu,m}, \text{ both } m, \mu \in (0, M), \tag{3}$$

then

$$W(mT) = \sum_{0}^{M} W_m(nT)\psi(nT), \qquad m \in (0, M), \tag{4}$$

or $W_\mu(nT)$ is μ, nth element of the inverse to the covariance matrix.

The conventional way of solving (1) or (3) has been to establish a square system of M equations, one for each required time point, and then to invert this system by algebraic techniques. In general, this requires the inversion of symmetrical matrices of rank equal to the memory M of the filter W.

Wise [5] and Siddiqui [6] have shown [7] that the problem of inverting such covariance matrices of rank M is reducible to the inversion of certain related matrices of rank $2D$, where $2D$ is the degree of the denominator of the power spectral density function. The former, using a semi-infinite matrix representation of translation operations instead of a spectral one, obtains a concise and elegant formal solution. However, but for a special case, the solution requires the inverse of semi-infinite triangular matrices. The latter treats the problem in statistical language, as a transformation of variables under a multivariate normal distribution. The symmetry and persymmetry properties of the covariance matrices are exploited to obtain a solution dependent on the inverse of a covariance matrix of rank $2D$.

The author independently derived similar conclusions about the equivalent problem of inverting (1). The required number of linearly independent equations to be solved equalled $2D$, where the memory M is greater than, or equal to, $2D$. But here, since (1) is of a time origin invariant form, the methods of time invariant harmonic analysis were used more extensively to obtain a finite, reduced triangular system of linear equations.

Thus, the purpose of the following investigation is to solve (3) by means of the discrete process harmonic analytical methods [8, 4], operationally referred to as the two-sided z transform [9]. The mode of approach in the transform domain is similar to that taken by Youla [10], although in the latter's article, continuous processes are dealt with in the solution of an eigenvalue problem.

Properties of Spectral Density Function

Since $\phi(mT)$, given by (2), is a combination of decaying exponentials, its two-sided z transform exists and is summable in closed form. It is given by

$$\Phi^*(z) = \sum_{-\infty}^{\infty} \phi(mT)z^{-m} = \sum_{k=1}^{D} C_k \frac{1 - e^{-2\alpha_k T}}{(1 - e^{-\alpha_k T}z)(1 - e^{-\alpha_k T}z^{-1})}, \tag{5}$$

for $-\alpha_k^- < \log|z| < \alpha_k^-$, hence for $z \neq 1$, or values of z on the unit circle (where α_k^- is the least of the α_k's).

Notice that the α_k's are assumed real in (5). If they are complex, then with each term shown on the right of (5) corresponding to an index k, there would have to be added a similar one, but with α_k replaced by α_k conjugate, since the correlation function is real-valued and even. Then the condition for convergence would be $-Re\,\alpha_k^- < \log|z| < Re\,\alpha_k^-$. The use of complex exponentials will only add to the complexity of a calculation without contributing anything of greater generality. Hence, only real coefficients will be used in the subsequent investigation. A summarizing discussion will deal with the appropriate interpretation of the results for the complex case.

The values that $\Phi^*(z)$ assumes on the unit circle $z = e^{i\omega T}$ are real and positive, and since

$$\oint_{(z\,|\,z=1)} |\log \Phi^*(z)| \frac{dz}{z} < \infty, \tag{6}$$

$\Phi^*(z)$ can be considered the analytic continuation of $\Phi(e^{i\omega T})$ within and without the unit circle. Further, from (5),

$$\Phi^*(z) = \Phi^*\left(\frac{1}{z}\right), \tag{7}$$

thus, $\Phi^*(z)$ is factorable [11]. In the particular case of (5) this follows by inspection.

$$\Phi^*(z) = \varphi(z)\varphi\left(\frac{1}{z}\right) \tag{8}$$

where $\varphi(z)$ is free of poles and zeros within the unit circle. Hence, the form for the inversion integrals,

$$\frac{1}{2\pi i}\oint_{(|z|=1)} \varphi(z)z^{m-1}\,dz, \tag{9}$$

is zero for all positive m, and

$$\frac{1}{2\pi i}\oint_{(|z|=1)} \varphi\left(\frac{1}{z}\right)z^{m-1}\,dz \tag{10}$$

is zero for all negative m.

Properties (6), (7), the positiveness of $\Phi(e^{i\omega T})$, and a theorem by Szego, allow the factorization (8). These properties and theorem are the discrete equivalent of the Paley-Wiener "factorability" criterion for continuous processes. Thus, from (5) and (8)

$$\Phi^*(z) = \frac{\varphi_N(z)\varphi_N\left(\frac{1}{z}\right)}{\varphi_D(z)\varphi_D\left(\frac{1}{z}\right)} \tag{11}$$

where $\varphi_D(z)$ is a polynomial of degree D, $\varphi_N(z)$ one of degree $N = D - P$, $P \geq 1$ and

$$\varphi(z) = \frac{\varphi_N(z)}{\varphi_D(z)} \tag{12}$$

occurs in (8).

Analysis

Let (3) be written as

$$v_\mu(m'T) = \sum_0^M W_\mu(mT)\phi((m' - m)T) - \delta_{\mu m'}; \tag{13}$$

then

$$v_\mu(m'T) = 0, \quad \text{both} \quad m, \mu \varepsilon (0, M). \tag{14}$$

Notice that outside the interval $(0, M)$, the first expression on the right of (13) "decays" in a manner determined by the poles $\phi(mT)$, since the convolved expression may be interpreted as the response of a nonrealizable digital filter (nonzero for both positive and negative time) to an input $W(mT)$ which is nonzero only over $(0, M)$.

It then follows that the z transform of (14) may be written as

$$V_\mu^*(z) = z^{-(M+1)}\left\{\frac{P\left(\frac{1}{z}\right)}{\varphi_D\left(\frac{1}{z}\right)}\right\} + z\left\{\frac{P'(z)}{\varphi_D(z)}\right\} \tag{15}$$

where $P_\mu(1/z)$ and $P'_\mu(z)$ are unknown polynomials of their respective arguments, each of degree not greater than $D - 1$ (since φ_D is of degree D), hence, each of not more than D coefficients.

The straightforward transformation of (13), since it holds for all m', gives us

$$V_\mu^*(z) = W^*(z)\Phi^*(z) - z^{-\mu}. \tag{16}$$

Thus, we may equate (15) with (16) and derive a more explicit form for $W_\mu^*(z)$:

$$W_\mu^*(z) = z^{-\mu}\frac{\varphi_D(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N(z)\varphi_N\left(\frac{1}{z}\right)} + \frac{z^{-(M+1)}P_\mu\left(\frac{1}{z}\right)\varphi_D(z) + zP'_\mu(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N(z)\varphi_N\left(\frac{1}{z}\right)}. \tag{17}$$

$W_\mu(nT)$ is a physically realizable weighting sequence of finite duration or memory. Hence, it is zero outside of $(0, M)$, which requires $W_\mu^*(z)$ to be a polynomial in z^{-1} of degree M. Thus, the problem is one of obtaining the coefficients of $P_\mu(1/z)$ and $P'_\mu(z)$, or a particular set of coefficients which are linearly and nonsingularly related to those of $P_\mu(1/z)$ and $P'_\mu(z)$ in order to satisfy the above condition.

If we expand (17) in its Dirichlet series form,

$$\begin{aligned}\sum_{m=0}^{M} z^{-m}W_\mu(mT) = \left\{\sum_{m=0}^{M} + \sum_{-\infty}^{-1} + \sum_{M+1}^{\infty}\right\}z^{-m} \\ \cdot\left\{\left[\frac{1}{\Phi(z)}\right][(m-\mu)T]\right. \\ + \left[\frac{P_\mu\left(\frac{1}{z}\right)\varphi_D(z)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}\right][(m-M-1)T] \\ \left. + \left[\frac{P'_\mu(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}\right][(m+1)T]\right\},\end{aligned} \tag{18}$$

where the large square brackets signify the time sequence whose transform is the enclosed expression.

Thus, for $W_\mu(z)$ to be an M degree polynomial in z^{-1}, it is both neccessary and sufficient that the terms summed as

$$\left\{\sum_{-\infty}^{-1} + \sum_{M+1}^{\infty}\right\}$$

be identically zero. As a time sequence, we have

$$\left[\frac{P_\mu\left(\frac{1}{z}\right)\varphi_D(z)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}\right][(m-M-1)T] + \left[\frac{P'_\mu(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}\right][(m+1)T]$$

$$= -\left[\frac{1}{\Phi(z)}\right][(m-\mu)T]$$

$$\text{for} \quad m \notin (0, M), \qquad \mu \varepsilon (0, M). \tag{19}$$

It is then necessary to obtain explicitly the expressions

$$\frac{P_\mu\left(\frac{1}{z}\right)\varphi_D(z)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)} \quad \text{and} \quad \frac{P'_\mu(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}$$

which satisfy (19). These functions are obtainable as linear combinations of known rational functions in z and z^{-1}. The coefficients of these known functions are then to be determined. $P_\mu(1/z)$ and $P'_\mu(z)$ are unknown polynomials of degree not greater than $D-1$; thus assume they both are of degree $D-1$, with possible zero coefficients. Hence,

$$\begin{aligned}\frac{P_\mu(\ )}{\varphi_N(\ )} &= Q_\mu(\ ) + \frac{R_\mu(\ )}{\varphi_N(\ )}, \\ \frac{P'_\mu(\ )}{\varphi_N(\ )} &= Q'_\mu(\ ) + \frac{R'_\mu(\ )}{\varphi_N(\ )},\end{aligned} \tag{20}$$

where Q_μ, Q'_μ are of degree $D-N-1$ and R_μ, R'_μ of degree $N-1$. Similarly,

$$\frac{\varphi_D(\ )}{\varphi_N(\ )} = \bar{Q}(\ ) + \frac{\bar{R}(\ )}{\varphi_N(\ )} \tag{21}$$

for $\bar{Q}$ and $\bar{R}$ of degree $D-N$ and $N-1$ respectively.

Thus, (19) becomes

$$\left[\left[Q_\mu\left(\frac{1}{z}\right) + \frac{R_\mu\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)\right][(m-M-1)T]$$

$$+\left[\left(Q'_\mu(z) + \frac{R_\mu(z)}{\varphi_N(z)}\right)\left[\bar{Q}\left(\frac{1}{z}\right) + \frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\right][(m+1)T]$$

$$= -\left[\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)\left[\bar{Q}\left(\frac{1}{z}\right) + \frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\right][(m-\mu)T] \tag{22}$$

for $m \notin (0, M)$, $\mu \in (0, M)$; thus, for m in the intervals $(M+1, M+D-N)$ and $(-1, -D+N)$, all of the terms of the above equations are necessary. This gives $2D-2N$ equations over the $2D-2N$ time points.

In consideration of the fact that

$$\oint dz\, z^{m-1}z^{-n} = 0, \qquad m \neq n,$$

when $m \geq M+1+D-N$ and $M \leq D-N-1$, the following modified equations are obtained:

a) $m \geq M+1+D-N$,

$$\left[\frac{R_\mu\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)\right][(m-M-1)T]$$

$$+\left[\left(Q'_\mu(z) + \frac{R'_\mu(z)}{\varphi_N(z)}\right)\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right][(m+1)T]$$

$$= -\left[\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right][(m-\mu)T], \tag{23}$$

b) $m \leq -D+N-2$,

$$\left[\left[Q_\mu\left(\frac{1}{z}\right) + \frac{R_\mu\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\frac{\bar{R}(z)}{\varphi_N(z)}\right][(m-M-1)T]$$

$$+\left[\frac{R'_\mu(z)}{\varphi_N(z)}\left[\bar{Q}\left(\frac{1}{z}\right) + \frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\right][(m+1)T]$$

$$= -\left[\left[\bar{Q}\left(\frac{1}{z}\right) + \frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\frac{\bar{R}(z)}{\varphi_N(z)}\right][(m-\mu)T]. \tag{24}$$

Let us choose our coefficients so that

a) (22) is true for $m \geq M+1$ and (25)

b) (23) is true for $m \leq -1$.

It is clear that (25) a) and b) imply (23) and (24), respectively, so there is no loss in generality. Thus, (22) is then reducible to two independent sets of equations as follows:

$$\left[\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)Q_\mu\left(\frac{1}{z}\right)\right][(m-M-1)T]$$

$$= \left[\left(\bar{Q}(z) + \frac{\bar{R}(z)}{\varphi_N(z)}\right)\bar{Q}\left(\frac{1}{z}\right)\right][(m-\mu)T],$$

$$0 \leq \mu \leq M, \quad M+1+D-N \geq m \geq M+1. \tag{26}$$

This admits a solution of the form

$$\left[Q_\mu\left(\frac{1}{z}\right)\right](nT) = \left[\bar{Q}\left(\frac{1}{z}\right)\right][(n+M+1-\mu)T]$$

$$n \geq 0. \tag{27}$$

Since both $Q_\mu(\ )$ and $\bar{Q}(\ )$ are $D-N-1$ degree polynomial forms this yields the $D-N$ coefficients of

$Q_\mu(\ )$ immediately. Similarly, for $-1 \leq m \leq -D+N$, (22) takes the form

$$\left[\left[\bar{Q}\left(\frac{1}{z}\right)+\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi\left(\frac{1}{z}\right)}\right]Q'_\mu(z)\right][(m+1)T]$$

$$=\left[\left[\bar{Q}\left(\frac{1}{z}\right)+\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right]\bar{Q}(z)\right][(m-\mu)T] \tag{28}$$

with the solution analogous to that of Q_μ,

$$[Q'_\mu(z)](nT) = [\bar{Q}(z)][(n-1-\mu)T]$$
$$n \leq 0. \tag{29}$$

Thus, $2(D-N)$ coefficients have been obtained. Note that if $N=0$, the problem is solved; this is the condition for which Wise's solution takes a particularly elegant form. The case for $N \pm 0$ will now be treated. The approach will be to use the already determined $2(D-N)$ coefficients of (27) and (29) in (23) and (24), express (23) and (24) in terms of arguments defined over $(0, \infty)$ and $(-1, \infty)$, respectively, and then to establish sets of difference equations over the respective domains by transposition of the denominators, $\varphi_N(1/z)$ and $\varphi_N(z)$.

Consider the definitions $\bar{R}_p$, $\bar{R}_{p'}$ such that

$$\left\{\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right\}[(p+n)T] = \left\{\frac{\bar{R}_p\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)}\right\}(nT);\ p,n \geq 0 \tag{30}$$

and

$$\left[\frac{\bar{R}(z)}{\varphi_N(z)}\right][(p+n)T] = \left[\frac{\bar{R}'_p(z)}{\varphi_N(z)}\right](nT);\ p,n \geq 0. \tag{31}$$

This is justified easily by the following. Let

$$\bar{R}\left(\frac{1}{z}\right) = \sum_0^{N-1} \bar{r}_n z^{-n} \tag{32}$$

and

$$\frac{\bar{R}\left(\frac{1}{z}\right)}{\varphi_N\left(\frac{1}{z}\right)} = \sum_1^N \frac{\bar{P}_K}{1-\gamma_K^{-1}z^{-1}}. \tag{33}$$

Then it follows that

$$\bar{R}_p\left(\frac{1}{z}\right) = \sum_0^{N-1} \bar{r}_n(p) z^{-n}, \tag{34}$$

where

$$\bar{r}_n(p) = (-)^n \sum_{K=1}^N \sum_{K'=1}^N \gamma_{K'}^{-p}\bar{\rho}_{K'} \prod_n \gamma_{k_1}^{-1} \cdots \gamma_{k_n}^{-1}$$
$$(k_1 \neq k_2 \neq \cdots \neq K_n \neq K'), \tag{35}$$

$\prod_n$ denoting the product over n different subscripts, n running from 0 to $N-1$. It is also true that

$$\bar{R}'_{-1p1}(z) = \bar{R}_p(z). \tag{36}$$

Thus, if (23) and (24) are to hold for $m \geq M+1$ and $m \leq -1$, respectively, it is sufficient that the following modified systems be satisfied:

$$\left[\varphi_D(z)R_\mu\left(\frac{1}{z}\right) + \bar{R}_{M+2}\left(\frac{1}{z}\right)R'_\mu(z) + \varphi_D(z)\bar{R}_{M+1-\mu}\left(\frac{1}{z}\right)\right.$$
$$\left. + \varphi_N(z)Q'_\mu(z)\bar{R}_{M+2}\left(\frac{1}{z}\right)\right](nT) = 0, \qquad n \geq 0 \tag{37}$$

and

$$\left[\bar{R}_{M+2}(z)R_\mu\left(\frac{1}{z}\right) + \varphi_D\left(\frac{1}{z}\right)R'_\mu(z) + \varphi_D\left(\frac{1}{z}\right)\bar{R}_{M+1+\mu}(z)\right.$$
$$\left. + \varphi_N\left(\frac{1}{z}\right)Q_\mu\left(\frac{1}{z}\right)\bar{R}_{M+2}(z)\right](nT) = 0 \qquad n \leq 0. \tag{38}$$

Here the coefficients of $R_\mu(1/z)$ and $R'_\mu(z)$, each numbering $N-1$, must be determined. Now, let us consider the Dirichlet series forms,

$$\varphi_D(z) = \sum_0^D \varphi_n z^n, \tag{39}$$

$$\varphi_N(z)Q'_\mu(z) = \sum_0^{D-1} \chi_n z^n, \tag{40}$$

$$\varphi_N\left(\frac{1}{z}\right)Q_\mu\left(\frac{1}{z}\right) = \sum_0^{D-1} \psi_n z^{-n}, \tag{41}$$

$$R_\mu\left(\frac{1}{z}\right) = \sum_0^{N-1} r_{\mu n} z^{-n}, \tag{42}$$

$$R'_\mu(z) = \sum_0^{N-1} r'_{\mu n} z^n, \tag{43}$$

and (33) and (36). Eqs. (37) and (38) thus become two triangular systems of $N-1$ equations each:

$$\sum_{m=n}^{N-1} \varphi_{m-n} r_{\mu,m} + \sum_{m=0}^{N-1-n} \bar{r}_{n+m}(M+2)r'_{\mu,m}$$
$$= -\sum_{m=n}^{N-1} \{\varphi_{m-n}\bar{r}_m(M+1-\mu) + \chi_{m-n}\bar{r}_m(M+2)\} \tag{44}$$

and

$$\sum_{m=0}^{N-1-n} \bar{r}_{n+m}(M+2)r_m + \sum_{m=n}^{N-1} \varphi_{m-n} r'_{\mu,m}$$
$$= -\sum_{m=n}^{N-1} \{\varphi_{m-n}\bar{r}_m(+1+\mu) + \psi_{m-n}\bar{r}_m(M+2)\}$$
$$\{n = 0, 1, \cdots, N-1\}. \tag{45}$$

By a succession of elementary operations (row and column multiplication and addition), all but the diagonal terms of the matrix of system (44), (45) can be eliminated, set to zero. Thus, the necessary and sufficient condition for the matrix of (45), (46) to be nonsingular is that all of its diagonal terms be nonzero.

If we start with $n = N - 1$ and work our way to $n = 0$, both (44) and (45) are easily solved as triangular systems. But our solutions will be in the form of pairs.

$$\varphi_{N-1-n}r_n + \bar{r}_{N-1}(M+2)r'_{N-1-n} = f_{1n} \tag{46}$$

and

$$\bar{r}_{N-1}(M+2)r_n + \varphi_n r'_{N-1-n} = f_{2N-1-n} \tag{47}$$

$$n = 0, 1. \cdots, N-1,$$

where f_{1n} and f_{2N-1-n} are the triangular solutions.

Hence, with the f_{1n}, f_{2N-1-n}'s obtained, the coefficients of $R_\mu(1/z)$ and $R'_\mu(z)$, r_n and r'_n are respectively derivable from each pair of simultaneous equations (46) and (47).

More Explicit Representation of Solution Form

Let us make the further definitions—

a) $\rho_{\mu r} = \{\prod_{k,k\neq r}(1 - e^{-(\beta_k-\beta_r)T})\}^{-1} \sum_{n=0}^{N-1} r_{\mu n}e^{-\beta_r T_n}$,

b) $\rho'_{\mu r} = \{\prod_{k,k\neq r}(1 - e^{-(\beta_k-\beta_r)T})\}^{-1} \sum_{n=0}^{N-1} r'_{\mu n}e^{-\beta_r T_n}$,

c) $\bar{Q}(z) = \sum_{0}^{D-N} \bar{Q}_r z^r$,

d) $Q_\mu(z) = \sum_{0}^{D-N-1} Q_{\mu r}z^r$

e) $Q'(z) = \sum_{0}^{D-N-1} Q'_{\mu r}z^r$

f) $q_r = \sum_{r'=1}^{D-N} \bar{Q}_{r'}\bar{Q}_{r'-r}$,

g) $q_0 = \sum_{0}^{D-N} \bar{Q}_r^2$,

h) $b_{rr'} = \bar{Q}_r\bar{\rho}'_r$,

i) $C_r = \sum_{r'=1}^{N} \frac{\bar{\rho}_r\bar{\rho}'_r}{1 - e^{-(\beta_r+\beta_{r'})T}}$,

j) $E_{\mu\bar{r}} = \sum_{r=0}^{r+D-N-1} \bar{Q}_r Q_{\mu,r-\bar{r}}$,

k) $F_\mu = \bar{Q}_0 Q_{\mu,1} + \bar{Q}_1 Q_{\mu,0}$,

l) $G_{\mu\bar{r}} = \sum_{D-N}^{\bar{r}} \bar{Q}_r Q_{\mu,r-\bar{r}}$,

m) $H_{\mu r} = \sum_{r''=1}^{N} \frac{\bar{\rho}_r\rho'_{\mu r} + \bar{\rho}'_r\rho_{\mu r}}{1 - e^{-(\beta_r+\beta_{r'})T}}$ and (48)

n) The primed coefficients E'_r, F', G'_r, and H'_r have the same form as the corresponding ones in j) through m), but Q_r, ρ_r are replaced by Q'_r, ρ'_r.

Upon using the definitions in (48) in the explicit form of the inverse to the covariance matrix, operationally expressed as

$$\{\Phi^{-1}\}_{\mu,m} \equiv W_\mu(mT) = \left[\frac{1}{\Phi(z)}\right][(m-\mu)T]$$
$$+ \left[\frac{p_\mu\left(\frac{1}{z}\right)\varphi_D(z)}{\varphi_N\left(\frac{1}{z}\right)\varphi_N(z)}\right][(m-M-1)T]$$
$$+ \left[\frac{p'_\mu(z)\varphi_D\left(\frac{1}{z}\right)}{\varphi_N(z)\varphi_N\left(\frac{1}{z}\right)}\right][(m+1)T]$$

for $0 \leq m, \mu \leq M$, and

$$= 0, \text{ otherwise,} \tag{49}$$
$$\{\Phi\}_{\mu,m} \equiv \phi((\mu - m)T),$$

we obtain a more explicit representation of the solution form:

$$W_\mu(mT) = \sum_{r=1}^{D-N} q_r(\delta_{m-\mu,-r} + \delta_{m-\mu,r}) + q_0\,\delta_{m-\mu,0}$$
$$+ \sum_{r=0}^{D-N}\sum_{r'=1}^{N} b_{rr'}(e(+)^{-\beta_{r'}T(m-\mu-r)} + e(-)^{\beta_{r'}T(m-\mu+r)})$$
$$+ \sum_{r=1}^{N} C_r e^{-\beta_r T|m-\mu|}$$
$$+ \sum_{\bar{r}=-D+N-1}^{0} (E_{\mu\bar{r}}\,\delta_{m-M-1,r} + E'_{\mu\bar{r}}\,\delta_{m+1,r}$$
$$+ F_\mu\,\delta_{m-M-1,1} + F'_\mu\,\delta_{m+1,-1})$$
$$+ \sum_{\bar{r}=1}^{D-N} (G_{\mu\bar{r}}\,\delta_{m-H-1,-r} + G'_{\mu\bar{r}}\,\delta_{m+1,-r})$$
$$+ \sum_{r=0}^{D-N}\sum_{r'=1}^{N} \bar{Q}_r\rho_{\mu r'}e(+)^{-\beta_{r'}T(m-M-1+r)}$$
$$+ \rho'_{\mu r'}e(-)^{\beta_{r'}T(m+1-r)}$$
$$+ \sum_{r'=1}^{N}\sum_{r=1}^{N} (H_{rr'} + H'_{rr'})e^{-\beta_r T|m-M-1|}, \tag{50}$$

where the (+) and (−) labels refer to functions which are nonzero only for positive and negative arguments, respectively.

Conclusion

Implicit in the preceding analysis is the condition that the memory M of the filter, expressed as the number of samples of input functions that are operated upon, must be equal to or greater than, the number of exponentials D in the input correlation function. On this basis, the optimal parameters are obtainable in the solution of a system of $2D$ linear equations. If this memory is actually less than D, the solution still may be obtained by the given method, but the number of equations is equal to the memory. This follows by virtue of the actual effective number of linearly independent exponentials in the input correlation function over the interval $(0, M)$; thus, if $M < D$ we may choose M of the exponentials to represent

he behavior of the input correlation function. Although he method described by the preceding work will apply to his case, it does not appear to have any advantages over he conventional time domain matrix inversion techniques ince the number of equations and unknowns is the same n both cases.

If the correlation function has exponentially damped eriodic components, its transform will have conjugate airs of poles and zeros within and outside of the unit ircle. Thus, (22) will keep the same form, but in the artial fraction expansion of the quotient of

$$\frac{\varphi_D(z)}{\varphi_N(z)},$$

oth $f_{r'}$, $B_{r'}$ and their conjugates will occur. The form of he analysis, however, will remain the same.

Let us multiply the transform of the optimal weighting equence (50) by T, the sampling period set $z = e^{sT}$ and hen take the limit as $T \to 0$, but restrict the memory uration MT to remain constant. The discrete process pproaches a continuous one, and the z-transformed xpression (50) becomes the Laplace transform of a finite nemory filter which operates continuously over the nterval $(0, MT)$. Thus, the transform of the optimal nite memory filter for a continuous input process should e obtainable as an asymptotic form of the discrete ase [12].

Example

As an example, let us consider the case where signal nd noise are stationary random. Assume that the signal nd noise correlation functions are given by

$$\sigma_M^2 e^{-\beta|nT|} \quad \text{and} \quad \sigma_N^2 e^{-\gamma|nT|}, \tag{51}$$

espectively. T is the sampling interval. For no cross orrelation, the input autocorrelation function is

$$\phi(nT) = \sigma_M^2 e^{-\beta|nT|} + \sigma_N^2 e^{-\gamma|nT|}. \tag{52}$$

The z transform of the sequence $\phi(nT)$, has the form

$$\phi^*(z) = \frac{\sqrt{k}\,(1 - e^{-\alpha T}z^{-1})}{(1 - e^{-\beta T}z^{-1})(1 - e^{-\gamma T}z^{-1})} \cdot \frac{\sqrt{k}\,(1 - e^{-\alpha T}z)}{(1 - e^{-\beta T}z)(1 - e^{-\gamma T}z)} = \frac{\varphi_N(z)\varphi_N\left(\frac{1}{z}\right)}{\varphi_D(z)\varphi_D\left(\frac{1}{z}\right)}, \tag{53}$$

where

$$k = \frac{2}{B_1 + \sqrt{B_1^2 - 4A_1^2}};$$

$$A_1 = \sigma_M^2(1 - e^{-2\beta T})e^{-\gamma T} + \sigma_N^2(1 - e^{-2\gamma T})e^{-\beta T}, \tag{54}$$

$$B_1 = (\sigma_M^2 + \sigma_N^2)(1 - e^{-2\beta T})(1 - e^{-2\gamma T}), \tag{55}$$

nd

$$\alpha = \frac{1}{T}\ln\left(\frac{B_1}{2A_1} + \sqrt{\frac{B_1^2}{4A_1^2} - 1}\right). \tag{56}$$

Here, it is assumed that $B_1 > 2A_1$. This does not restrict the subsequent procedure except in eliminating the temporary introduction of complex quantities. Thus,

$$\frac{\varphi_D(z)}{\varphi_N(z)} = \bar{Q}_1 z + \bar{Q}_0 + \frac{\bar{p}_1}{1 - e^{-\alpha T}z}, \tag{57}$$

where

$$\bar{Q}_1 = \frac{-e^{-(\beta+\gamma+\alpha)T}}{\sqrt{k}},$$

$$\bar{Q}_0 = \frac{-1}{\sqrt{k}}(e^{-(\beta+\gamma-2\alpha)T} - e^{-(\beta-\alpha)T} - e^{-(\gamma-\alpha)T})$$

$$\bar{p}_1 = \frac{-1}{\sqrt{k}}(1 + e^{-(\beta+\gamma-2\alpha)T} - e^{-(\beta-\alpha)T} - e^{-(\gamma-\alpha)T}). \tag{58}$$

Here we have $D = 2$, $N = 1$, $D - N = 1$. Thus, $\bar{Q}(1/z) = \bar{Q}_0 + \bar{Q}_1 z^{-1}$ and from (27) and (29), respectively,

$$\left[Q_\mu\left(\frac{1}{z}\right)\right](nT) = \left[\bar{Q}\left(\frac{1}{z}\right)\right][(n + M + 1 - \mu)T] = \bar{Q}_1\,\delta_{n+M+1-\mu,1} = \bar{Q}_1\,\delta_{n,0}\,\delta_{\mu,M} \tag{59}$$

$$[Q'_\mu(z)](nT) = [\bar{Q}(z)][(n - 1 - \mu)T] = \bar{Q}_1\,\delta_{n-1-\mu,-1} = \bar{Q}_1\,\delta_{n,0}\,\delta_{\mu,0}. \tag{60}$$

Upon using (30), or (35),

$$\bar{R}_\nu\left(\frac{1}{z}\right) = \bar{R}_\nu(z) = e^{-\alpha\nu T}\bar{p}_1. \tag{61}$$

Since $N - 1 = 0$, only the zeroth terms in (39) through (45) are necessary. Note that [from (39), (40), and (41), respectively],

$$\varphi_0 = 1, \qquad x_0 = \bar{Q}_1\,\delta_{\mu,M}, \qquad \Psi_0 = \bar{Q}_1\,\delta_{\mu,0}. \tag{62}$$

Hence, the resulting pair of equations,

$$r_{\mu,0} + \bar{p}_1 e^{-\alpha T(M+2)} r'_{\mu,0} = \bar{p}_1 e^{-\alpha T(M+1-\mu)} + \bar{p}_1\bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,M}$$

and

$$\bar{p}_1 e^{-\alpha T(M+2)} r_{\mu,0} + r'_{\mu,0} = \bar{p}_1 e^{-\alpha T(1+\mu)} + \bar{p}_1\bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,0}. \tag{63}$$

Thus, the remaining unknowns $r_{\mu,0}$ and $r'_{\mu,0}$ are easily obtained:

$$r_{\mu,0} = \{1 - \bar{p}_1^2 e^{-2\alpha T(M+2)}\}^{-1} \cdot \{\bar{p}_1(e^{-\alpha T(M+1-\mu)} + \bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,M}) - \bar{p}_1^2(e^{-\alpha T(1+\mu)} + \bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,0})\}$$

and

$$r'_{\mu,0} = \{1 - \bar{p}_1^2 e^{-2\alpha T(M+2)}\}^{-1} \cdot \{\bar{p}_1(e^{-\alpha T(1+\mu)} + \bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,0}) - \bar{p}_1^2(e^{-\alpha T(n+1-\mu)} + \bar{Q}_1 e^{-\alpha T(M+2)}\,\delta_{\mu,M})\}. \tag{64}$$

Hence, from (59) and (60), referring to (48d) and (48e),

$$Q_{\mu,r} = \bar{q}_1 \, \delta_{r,0} \, \delta_{\mu,M} \quad \text{and} \quad Q'_{\mu,r} = \bar{q}_1 \, \delta_{r,0} \, \delta_{\mu,0}, \tag{65}$$

and the four coefficients $Q_{\mu,0}$, $Q'_{\mu,0}$, $r_{\mu,0}$, $r'_{\mu,0}$ have been obtained. Since the degree of φ_D is 2, the necessary and sufficient number of coefficients is also $2D = 4$.

On referring to definitions (48), the μ, mth element of the inverse matrix takes the form

$$\begin{aligned}
W_\mu(mT) = {} & (\bar{Q}_0^2 + \bar{Q}_1^2) \, \delta_{m-\mu,0} \\
& + 2\bar{Q}_1\bar{Q}_0(\delta_{m-\mu,-1} + \delta_{m-\mu,1}) \\
& + \bar{Q}_1 p_1 e(+)^{-\alpha T(m-\mu-1)} + e(-)^{-\alpha T(m-\mu+1)} \\
& + \frac{\bar{p}_1 2}{1 - e^{-2\alpha T}} \, e^{-\alpha T|m-\mu|} \\
& + \bar{Q}_0 Q_{\mu,0} \, \delta_{m-M-1,r} + \bar{Q}_0 Q'_{\mu,0} \, \delta_{m+1,r} \\
& + \bar{Q}_1 Q_{\mu,0} \, \delta_{m-M-1,1} + \bar{Q}_1 Q'_{\mu,0} \, \delta_{m+1,-1} \\
& + \bar{Q}_1 Q_{\mu,0} \, \delta_{m-M-1,-1} + \bar{Q}_1 Q'_{\mu,0} \, \delta_{m+1,-r} \\
& + r_{\mu,0}(\bar{Q}_0 e(+)^{-\alpha T(m-M-1)} + \bar{Q}_1 e(+)^{-\alpha(mM+1)}) \\
& + r'_{\mu,0}(\bar{Q}_0 e(-)^{\alpha T(m+1)} + \bar{Q}_1 e^{\alpha T m}) \\
& + \frac{\bar{p}_1(r'_{\mu,0} + r_{\mu,0})}{1 - e^{-2\alpha T}} \, e^{-\alpha T|m-M-1|}.
\end{aligned} \tag{66}$$

Bibliography

[1] A. C. Aitken, "On least square and linear combination of pr diction," *Proc. Roy. Soc. (Edinburgh)*, vol. 55, pp. 42–4 November, 1934.

[2] A. B. Lees, "Interpolation and extrapolation of sampled-data, IRE Trans. on Information Theory, vol. IT-2, pp. 12–1 March, 1956.

[3] M. Blum, "An extension of the minimum mean square pr diction theory for sampled input signals," IRE Trans. o Information Theory, vol. IT-2, pp. 176–184; Septembe 1956.

[4] N. Wiener, "Extrapolation, Interpolation and Smoothing Stationary Time Series," John Wiley and Sons, Inc., New York N. Y.; 1949.

[5] J. Wise, "The autocorrelation function and the spectral densit function," *Biometrika*, vol. 42, pp. 151–159; 1955.

[6] M. M. Siddiqui, "The inversion of the sample covarianc matrix in a stationary autoregressive process," *Annals of Math Statistics*, vol. 129, pp. 585–588; June, 1958.

[7] The author is indebted to M. Blum for this realization.

[8] H. Wold, "A Study in the Analysis of Stationary Time Series," Almquist and Wiksell, Stockholm, Sweden; 1938, 1954.

[9] G. Franklin, "Linear filtering of sampled-data," 1955 IRE Convention Record, pt. 4; pp. 119–128.

[10] D. C. Youla, "The solution of a homogeneous Wiener-Hop integral equation occurring in the expansion of second-orde stationary random functions," IRE Trans. on Information Theory, vol. IT-3, pp. 187–193; September, 1957.

[11] G. Szego, "On the boundary value of an analytic function," *Math. Ann.*, vol. 84, pp. 232–244; 1921.

[12] That the optimal continuous filter is attainable as an asymp totic form of the optimal digital filter has been proved in th time domain in the paper by P. Swerling, "Optimum linea estimation for random processes as the limit of estimate based on sampled data," 1958 IRE WESCON Convention pt. 4; pp. 158–163.

[13] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener' theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645–655 July, 1950.

[14] W. A. Janos, "Optimal Filtering of Periodic Pulse-Modulate Time Series," Ph.D. dissertation, University of California Berkeley, Calif.; 1958.

[15] R. Mittra, "On the Solution of a Class of Wiener-Hopf Integra Equations in Finite and Infinite Ranges," Antenna Lab., Uni versity of Illinois, Urbana, Ill., Tech. Rept. No. 37; 1959.

An Isospectral Family of Random Processes*

RICHARD A. SILVERMAN†, SENIOR MEMBER, IRE

Summary—We construct a family of random step functions $x_n(t)\}$ whose members all have the same power spectrum and ich that as $n \to \infty$, $x_n(t)$ converges to $x_\infty(t)$, the Gaussian process ith the same spectrum. We illustrate the procedure for calculating ie general multivariate distribution of the processes $\{x_n(t)\}$ by lculating the univariate, bivariate and trivariate distributions. Ve show how a suitably constructed univariate entropy can serve s an index of the extent to which $x_n(t)$ has approached the Gaussian mit $x_\infty(t)$.

Introduction

CONSIDER a zero-mean stationary random process $x(t)$ with correlation function $C(\tau)$, *i.e.*, a process $x(t)$ for which

$$Ex(t) = 0, \qquad Ex(t)\, x(t+\tau) = C(\tau),$$

here E is the expectation operator. A considerable part f noise theory is devoted to study of the case where $x(t)$ Gaussian. In this case, the *law* of $x(t)$, *i.e.*, the probbility that $x(t_n) < c_n$, $1 \leq n \leq N$, $1 \leq N < \infty$, where he c_n are arbitrary real numbers, is the familiar N-variate aussian distribution whose moment matrix can be xpressed simply in terms of $C(\tau)$.[1] Thus, in the Gaussian ase, $C(\tau)$ uniquely specifies $x(t)$, which is, of course, the articular beauty of this case. More generally, $C(\tau)$ gives nly a more or less incomplete characterization of $x(t)$.

The object of the present paper is to describe and study n infinite family $\{x_n(t)\}$, $1 \leq n < \infty$, of non-Gaussian andom processes, which converge to a Gaussian process $_\infty(t)$ as $n \to \infty$, and *whose laws are calculable, at least in rinciple*. The family $\{x_n(t)\}$ will be constructed in such way that all its members have the same correlation unction $C(\tau)$, or equivalently, the same power spectrum

$$\Phi(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega\tau) C(\tau)\, d\tau.$$

Briefly, we shall say that $\{x_n(t)\}$ is an *isospectral* family. ince as $n \to \infty$, the general appearance of the sample unctions of $x_n(t)$ changes and approaches that of the aussian process $x_\infty(t)$ with the same power spectrum or correlation function), the general inadequacy of (ω) or $C(\tau)$ as a means of characterizing random processes and the general need for higher-order statistics are put into rather strong focus.

The method we use to construct $x_n(t)$ from an underlying shot noise (Poisson process) is a familiar one. The novelty of our treatment consists in observing that when the shots are rectangular the members of $\{x_n(t)\}$ can be described in theoretically complete detail.

Construction of $\{x_n(t)\}$

Let the sequence t_j ($j = \cdots, -1, 0, 1, \cdots$) be the occurrence times of the shots in a stationary shot noise (Poisson process) with an average rate of ρ shots per second. More specifically, we have the following five properties (among others):

1) The occurrence of m shots in the interval I and the occurrence of n shots in the interval I' ($m, n = 0, 1, 2, \cdots$) are independent events if I and I' do not overlap.

2) The probability of one shot in an infinitesimal interval of length Δt is $\rho\Delta t + o(\Delta t)$, whereas the probability of more than one shot is $o(\Delta t)$.

3) The probability of m shots in any interval of length T is given by the Poisson distribution

$$p(m; \lambda) = \frac{\lambda^m}{m!} e^{-\lambda}, \qquad m = 0, 1, 2, \cdots,$$

where $\lambda = \rho T$, the parameter of the Poisson distribution, is the common value of the mean and variance of a random variable which takes the values $m = 0, 1, 2, \cdots$ with probabilities $p(m; \lambda)$.

4) Given that m shots have occurred in an interval of length T, their occurrence times $t_1, \cdots, t_m$ (without regard to order of appearance) are independent, identically distributed random variables with the uniform probability density $1/T$.

5) The probability density of the interval between successive shots is $\rho \exp(-\rho t)$.

For discussion and derivation of these properties we refer the reader elsewhere.[2-4]

Now let $h(t; \alpha)$ be a step function of unit height and width α, *i.e.*,

$$\begin{aligned} h(t;\alpha) &= 1, \qquad 0 \leq t \leq \alpha, \\ h(t;\alpha) &= 0, \qquad -\infty < t < 0, \qquad \alpha < t < \infty. \end{aligned} \tag{1}$$

* Received by the PGIT, January 11, 1960. The research reorted in this article has been sponsored by the Air Force Camridge Research Center, Air Research and Development Command, nder Contract No. AF 19(604)5238.

† Institute of Mathematical Sciences, New York University, 5 Waverly Place, New York, N. Y.

[1] S. O. Rice, "Mathematical Analysis of Random Noise," reprinted in the collection "Selected Papers on Noise and Stochastic rocesses," N. Wax, Ed., Dover Publications, Inc., New York, . Y., pp. 181–183; 1954.

[2] W. Feller, "An Introduction to Probability Theory and its Applications," 2nd ed., John Wiley and Sons, Inc., New York, N. Y., pp. 146, 400; 1957.

[3] Rice, *op. cit.*, see pt. I.

[4] A. Blanc-Lapierre and R. Fortet, "Théorie des Fonctions Aléatoires," Masson, Paris, ch. 5; 1953.

To construct the isospectral family $\{x_n(t)\}$, we write

$$x_n(t) = (1/\sqrt{n}) \sum_{j=-\infty}^{\infty} h(t - t_j^{(n)}; \alpha) - \sqrt{n}\alpha,$$

$$1 \leq n < \infty, \qquad (2)$$

where the $t_j^{(n)}$ are random times belonging to a shot noise with an average rate of n shots per second. It is clear that each $x_n(t)$ is a stationary random step function. To calculate expectations of lagged products of $x_n(t)$ we follow the usual procedure: First we use 4) to calculate conditional expectations for a long finite interval of length T containing exactly N shots and then we average over N using 3). Again, we refer to the literature for details.[5] In particular, we find

$$Ex_n(t) = 0, \qquad Ex_n(t)x_n(t + \tau) = C(\tau), \qquad 1 \leq n < \infty,$$

where $C(\tau)$ is given by

$$C(\tau) = \int_{-\infty}^{\infty} h(t; \alpha)h(t + \tau; \alpha)\, dt$$

$$= \alpha - |\tau|, \qquad 0 \leq |\tau| \leq \alpha, \qquad (3)$$

$$C(\tau) = 0, \qquad |\tau| > \alpha.$$

The corresponding common power spectrum of the isospectral family $\{x_n(t)\}$ is

$$\Phi(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\omega\tau)C(\tau) = (2/\pi\omega^2) \sin^2(\omega\alpha/2).$$

The constant α is still at our disposal. In order to make the members of $\{x_n(t)\}$ corresponding to small values of n drastically non-Gaussian, we can choose $\alpha \ll 1$, so that there is negligible overlap of the shots making up $x_n(t)$ if $n \ll 1/\alpha$. Then for $n \sim 1/\alpha$ overlap of the shots begins to be appreciable and as n increases further, the process $x_n(t)$ approaches the Gaussian process $x_\infty(t)$ with the same power spectrum.[6]

Typical Sample Functions of $\{x_n(t)\}$

To construct typical sample functions of $x_n(t)$, we first use random number tables[7] to find sample values of a random variable γ which is uniformly distributed in the unit interval (0, 1). We then observe that the new random variable

$$\eta = \frac{1}{\rho} \log \frac{1}{1 - \gamma} \qquad (4)$$

has the probability density $\rho \exp(-\rho t)$ of the interval between successive shots in a Poisson process with averag rate ρ. To see this, note that

$$\text{Prob}\,[F^{-1}(\gamma) < t] = \text{Prob}\,[\gamma < F(t)] = F(t),$$

so that $F^{-1}(\gamma)$ has the distribution function $F(t)$. If $F($ is to be the distribution function corresponding to th probability density which is $\rho \exp(-\rho t)$ for $t \geq 0$ ar zero otherwise, then $F(t) = 1 - \exp(-\rho t)$, whence (follows at once. Moreover, if γ is uniformly distributed (0, 1) so is $1 - \gamma$. Thus, sample values of the rando variable.

$$\eta_n = \frac{1}{n} \log \frac{1}{\gamma} \qquad (5$$

generate time markers locating the shots in a Poisso process with an average rate of ρ shots per second. Speci cally, if $\eta_{n1}, \cdots, \eta_{nm}$ are m sample values of (5), we g m random times $t_1^{(n)}, \cdots, t_m^{(n)}$ by writing

$$t_1^{(n)} = \eta_{n1}, \qquad t_2^{(n)} = \eta_{n1} + \eta_{n2}, \cdots,$$

$$t_m^{(n)} = \eta_{n1} + \eta_{n2} + \cdots \eta_{n}$$

Finally, with these values of the random times, we u (2) to construct sample functions of $x_n(t)$.

Figs. 1–3 show the results of this procedure for th cases of low, medium and high density shot noise, r spectively. In each case α was chosen to be 0.25. Fig. shows a four-second sample of the process

$$x_1(t) = \sum_{j=-\infty}^{\infty} h(t - t_j^{(1)}; 0.25) - 0.25,$$

Fig. 2 shows a four-second sample of the process

$$x_4(t) = \tfrac{1}{2} \sum_{j=-\infty}^{\infty} h(t - t_j^{(4)}; 0.25) - 0.50,$$

and Fig. 3 shows a one-second sample (drawn on a differe scale) of the process

$$x_{16}(t) = \tfrac{1}{4} \sum_{j=-\infty}^{\infty} h(t - t_j^{(16)}; 0.25) - 1.00.$$

The circle in Fig. 2 shows a level where the sojourn $x_4(t)$ was so brief that it could not be indicated on th scale of Fig. 2. A similar remark applies to the thr circles appearing in Fig. 3. In each case, the sample fun tions were given the initial value of $-\sqrt{n}\alpha$, *i.e.*, $-0.$ for $x_1(t)$, -0.50 for $x_4(t)$ and -1.00 for $x_{16}(t)$. Wh starting the sample functions of $x_4(t)$ and $x_{16}(t)$, atypic portions lasting about $\alpha = 0.25$ seconds occur, due the fact that there are no shots before $t = 0$.[8] (The retically, each $x_n(t)$ should begin in the infinitely remo past.) To eliminate this transient behavior, we suppress the first second of the record of $x_4(t)$ and the first second of the record of $x_{16}(t)$.

[5] Rice, *op. cit.*, see pts. I, II.

[6] Specifically, the type of convergence we have in mind is convergence in distribution, *i.e.*, the law of $x_n(t)$ converges to that of $x_\infty(t)$.

[7] The RAND Corporation, "A Million Random Digits with 100,000 Normal Deviates," Free Press, Glencoe, Ill.; 1955.

[8] This effect is unimportant for $x_1(t)$, since in this case the ov lap of shots is slight.

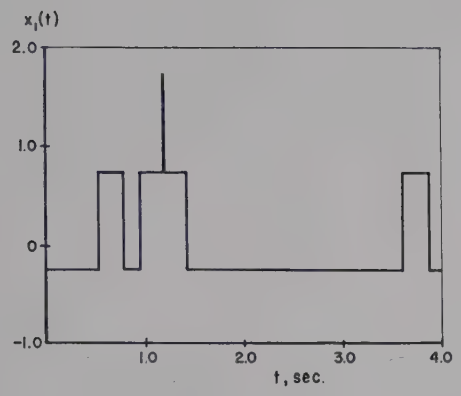


Fig. 1—Sample function of $x_1(t)$.

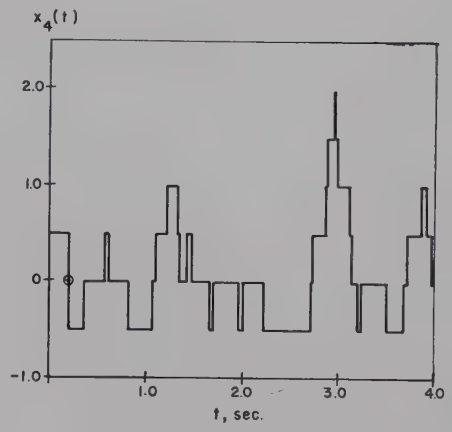


Fig. 2—Sample function of $x_4(t)$.

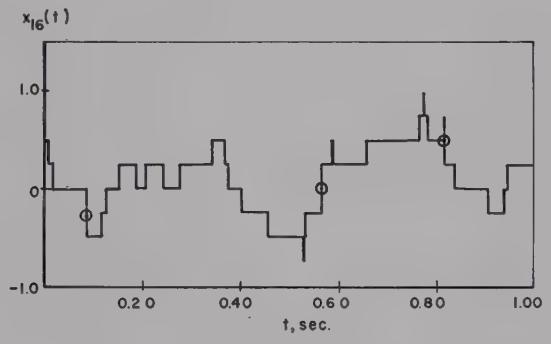


Fig. 3—Sample function of $x_{16}(t)$.

Probability Distributions of $\{x_n(t)\}$

The family $\{x_n(t)\}$ has the desirable feature that we can calculate the law of each $x_n(t)$, *i.e.*, all the multivariate probability distributions of $x_n(t)$, although the amount of work required to calculate high-order probability distributions is considerable. We shall illustrate the general procedure by deriving formulas for the univariate, bivariate and trivariate probability distributions of $x_n(t)$.

A. Univariate Distribution

By (1) and (2), the range of possible values of the random variable $x_n(t)$, t fixed,[9] is the lattice

$$y_{m,n} = m/\sqrt{n} - \sqrt{n}\alpha, \qquad m = 0, 1, 2, \cdots .$$

The probability that $x_n = y_{m,n}$ is just the probability that in the α seconds preceding t precisely m shots occur. Consequently, denoting the probability that $x_n = y_{m,n}$ by $P_n(m)$, we have

$$P_n(m) = p(m; n\alpha) = \frac{(n\alpha)^m}{m!} e^{-n\alpha}, \tag{6}$$

where $p(m; n\alpha)$ is the Poisson distribution with parameter $n\alpha$. Elementary calculations verify that

$$\sum_{m=0}^{\infty} y_{m,n} P_n(m) = 0, \qquad \sum_{m=0}^{\infty} y_{m,n}^2 P_n(m) = \alpha, \quad 1 \leq n \leq \infty,$$

as required. Note that $P_n(m)$ can be regarded as the sum of n independent, identically distributed Poisson random variables with parameter α. Thus, the convergence of $x_n(t)$ to a zero-mean Gaussian random variable with variance α is a particularly simple case of the central limit theorem.[10] Note also that $x_n(t)$ can take arbitrarily large positive values, but no negative values less than $-\sqrt{n}\alpha$; this asymmetry of the sample functions disappears as $n \to \infty$.

B. Bivariate Distribution

The probability that $x_n(t) = y_{\ell,n}$ while $x_n(t + \tau) = y_{m,n}$ will be denoted by $P_n(\ell, m; \tau)$ and is independent of t, since $x_n(t)$ is stationary. The event that precisely m shots occur in the interval (u, v) will be denoted by $E_m(u, v)$. Then, clearly, $P_n(\ell, m; \tau)$ is the joint probability

[9] The value of t is irrelevant since $x_n(t)$ is stationary, so that we can drop the argument t.

[10] The fact that the Poisson distribution $p(m; \lambda)$ approximates the Gaussian distribution for large values of λ is noted by Feller, *op. cit.*, see p. 176.

of the events $E_\ell\ (-\alpha, 0)$ and $E_m\ (\tau - \alpha, \tau)$. If $|\tau| \geq \alpha$, these two events are independent and we have

$$P_n(\ell, m; \tau) = P_n(\ell)P_n(m) = \frac{(n\alpha)^\ell}{\ell!}\frac{(n\alpha)^m}{m!}e^{-2n\alpha}.$$

On the other hand, if $\tau = 0$, we obviously have

$$P_n(\ell, m; 0) = P_n(\ell)\ \delta_{\ell m},$$

where $\delta_{\ell m}$ is the Kronecker delta. The interesting case is $0 < |\tau| < \alpha$, for then $E_\ell\ (-\alpha, 0)$ and $E_m\ (\tau - \alpha, \tau)$ are not independent events, since the intervals $(-\alpha, 0)$ and $(\tau - \alpha, \tau)$ overlap. In this case, we have

$$P_n(\ell, m; \tau) = \sum \text{Prob}\ \{E_{\ell-s}(-\alpha, \tau - \alpha), E_s(\tau - \alpha, 0), E_{m-s}(0, \tau)\}$$

if $0 < \tau < \alpha$, and

$$P_n(\ell, m; \tau) = \sum \text{Prob}\ \{E_{m-s}(\tau - \alpha, -\alpha), E_s(-\alpha, \tau), E_{\ell-s}(\tau, 0)\}$$

if $-\alpha < \tau < 0$; the summations are over all values of the nonnegative integer s compatible with the given values of ℓ and m. In either case, we find

$$P_n(\ell, m; \tau) = \sum_{s=0}^{\min(\ell,m)} \frac{(n\,|\,\tau\,|)^{\ell-s}(n(\alpha - |\,\tau\,|))^s(n\,|\,\tau\,|)^{m-s}}{(\ell - s)!\,s!\,(m - s)!} \cdot \exp(-n\,|\,\tau\,| - n\alpha), \qquad 0 < |\,\tau\,| < \alpha. \tag{7}$$

Eq. (7) is the explicit form of the bivariate Poisson distribution, defined by Feller[11] in terms of its generating function

$$E\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty} s_1^\ell s_2^m P_n(\ell, m; \tau) = \exp n[s_1s_2(\alpha - \tau) + s_1\tau + s_2\tau - \tau - \alpha], \qquad 0 < \tau < \alpha. \tag{8}$$

The generating function (8) can either be calculated directly (as pointed out by a referee) or by making the change of variables $s_1 = e^{iu}$, $s_2 = e^{iv}$ in the expression for the two-dimensional characteristic function

$$E\exp[iux_n(t) + ivx_n(t + \tau)] = \exp\left\{n\int_{-\infty}^{\infty}[e^{iuh(t;\alpha)+ivh(t+\tau;\alpha)} - 1]\right\}dt$$

$$= \exp\{n[\tau(e^{iv} - 1) + (\alpha - \tau)(e^{iu+iv} - 1) + \tau(e^{iu} - 1)]\}$$

given by Rice.[12] (In writing the last expression, we have simplified (2) by setting the normalization factor $1/\sqrt{n} = 1$ and dropping the centering constant $-\sqrt{n}\alpha$. Then, $P_n(\ell, m; \tau)$ means the joint probability that $x_n(t) = \ell$, $x_n(t + \tau) = m$ instead of the joint probability that $x_n(t) = y_{\ell,n}$, $x_n(t + \tau) = y_{m,n}$.)

Since $C(\tau)$, as given by (3), must be the correlation function of every $x_n(t)$, it follows that

$$\sum_{\ell,m=0}^{\infty} y_{\ell,n}y_{m,n}P_n(\ell, m; \tau) = C(\tau), \qquad 1 \leq n < \infty. \tag{9}$$

For $\tau = 0$ or $|\tau| \geq \alpha$, it is trivial to verify (9) directly but for $0 < |\tau| < \alpha$, (9) becomes highly nontransparent. A typical identity derivable from (9) is

$$\sum_{\ell,m=0}^{\infty}\frac{\ell m}{2^{\ell+m}}\sum_{s=0}^{\min(\ell,m)}\frac{2^s}{(\ell - s)!\,s!\,(m - s)!} = \tfrac{3}{2}\exp(3/2).$$

(Set $n = \alpha = 1$, $\tau = 1/2$.)

C. Trivariate Distribution

It is clear by now how the construction proceeds in general. Therefore, we calculate only the probability $P_n(k, \ell, m; \tau, \tau')$ of the joint event that $x_n(t) = y_{k,n}$, $x_n(t + \tau) = y_{\ell,n}$ and $x_n(t + \tau + \tau') = y_{m,n}$, assuming that $0 < \tau + \tau' < \alpha$. In this case, we have

$$\begin{aligned}P_n(k, \ell, m; \tau, \tau') &= \text{Prob}\ \{E_k(-\alpha, 0), E_\ell(\tau - \alpha, \tau), E_m(\tau + \tau' - \alpha, \tau + \tau')\} \\ &= \sum \text{Prob}\ \{E_{k-r-s}(-\alpha, \tau - \alpha), E_r(\tau - \alpha, \tau + \tau' - \alpha), E_s(\tau + \tau' - \alpha, 0), \\ &\qquad E_{\ell-r-s}(0, \tau), E_{m-\ell+r}(\tau, \tau + \tau')\},\end{aligned}$$

where we have gone over to nonoverlapping events, and the summation is over all values of the nonnegative integers r and s, compatible with the given values of k, ℓ and m. Specifically, we find

$$P_n(k, \ell, m; \tau, \tau') = \sum_{s=0}^{\min(k,\ell,m)}\sum_{r=-\min(m-\ell,0)}^{\min(k-s,\ell-s)}\frac{(n\tau)^{k-r-s}}{(k - r - s)!}\cdot\frac{(n\tau')^r}{r!}\frac{[n(\alpha - \tau - \tau'])^s}{s!}\frac{(n\tau)^{\ell-r-s}}{(\ell - r - s)!}\frac{(n\tau')^{m-\ell+r}}{(m - \ell + r)!}\cdot\exp(-n\tau - n\tau' - n\alpha).$$

For the three-dimensional characteristic function, we have (with the same definition of $x_n(t)$ as in the two-dimensional case)

$$\begin{aligned}&E\exp[iux_n(t) + ivx_n(t + \tau) + iwx_n(t + \tau + \tau')] \\ &= \exp\left\{n\int_{-\infty}^{\infty}[e^{iuh(t;\alpha)+ivh(t+\tau;\alpha)+iwh(t+\tau+\tau';\alpha)} - 1]\,dt\right. \\ &= \exp\{n[\tau'(e^{iw} - 1) + \tau(e^{iv+iw} - 1) \\ &\quad + (\alpha - \tau - \tau')(e^{iu+iv+iw} - 1) \\ &\quad + \tau'(e^{iu+iv} - 1) + \tau(e^{iu} - 1)]\}.\end{aligned}$$

As before, the generating function can be obtained from the characteristic function by a simple change of variable and is found to be[13]

$$\sum_{k,\ell,m=0}^{\infty} s_1^k s_2^\ell s_3^m P_n(k, \ell, m; \tau, \tau') = \exp\{n[s_1s_2s_3(\alpha - \tau - \tau' + s_1s_2\tau' + s_2s_3\tau + s_1\tau + s_3\tau' - \tau - \tau' - \alpha]\}$$

[11] Feller, *op. cit.*, see p. 261.
[12] Rice, *op. cit.*, see p. 245.
[13] Inclusion of these remarks concerning characteristic functions and generating functions was suggested by S. O. Rice.

The examples just given make it clear how to calculate $P_n(k, \ell, m; \tau, \tau')$ for other ranges of the parameters τ, τ' and how to calculate the general N-variate distribution $P_n(k_1, \cdots, k_N; \tau_1, \cdots, \tau_{N-1})$. Of course, for large N, explicit calculation of the N-variate distribution is formidable and would require the services of a high-speed electronic computer.

Entropy of $\{x_n(t)\}$

The fact that as $n \to \infty$, $x_n(t)$ converges in distribution to the Gaussian process $x_\infty(t)$ with the same power spectrum follows from the work of Fortet.[14] An index of the closeness of $x_n(t)$ to $x_\infty(t)$ is furnished by the entropy of $x_n(t)$; for simplicity, we consider only the univariate entropy of $x_n(t)$. Since the entropy of the Gaussian random variable x_∞ with mean zero and variance α, as defined in the usual way,[15] is

$$H_\infty(\alpha) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\alpha}} e^{-\xi^2/2\alpha} \log\left(\frac{1}{\sqrt{2\pi\alpha}} e^{-\xi^2/2\alpha}\right) d\xi$$

$$= \log \sqrt{2\pi e\alpha},$$

we must insist that the univariate entropy of x_n converge to $H_\infty(\alpha)$ as $n \to \infty$.[16] As we have seen, the univariate distribution of x_n is given by

$$P_n(m) \equiv \text{Prob}\,\{x_n = y_{m,n} \equiv m/\sqrt{n} - \sqrt{n}\alpha\}$$

$$= p(m; n\alpha) \equiv \frac{(n\alpha)^m}{m!} e^{-n\alpha}.$$

Thus, at first it might seem that the appropriate entropy to consider is the entropy

$$H_n(\alpha) = -\sum_{m=0}^{\infty} p(m; n\alpha) \log p(m; n\alpha) \qquad (10)$$

of the discrete random variable x_n. However, the limit as $n \to \infty$ of $H_n(\alpha)$, as defined by (10), is independent of α, since α is involved only in the combination $n\alpha$, and therefore, $H_n(\alpha)$ cannot converge to $H_\infty(\alpha) = \log \sqrt{2\pi e\alpha}$. In fact, $H_n(\alpha)$ is actually logarithmically divergent, as the following simple qualitative argument shows: As already noted, as $n \to \infty$, the random variable x_n converges in distribution to the Gaussian random variable x_∞ with mean zero and variance α. It follows that for large n most of the distribution of x_n is concentrated in the interval $(-\sqrt{\alpha}, +\sqrt{\alpha})$, *i.e.*, within one standard deviation of zero, and that the values of $p(m; n\alpha)$ are approximately equal in this interval. Since the possible values of x_n are a distance $1/\sqrt{n}$ apart, there are approximately $2\sqrt{n\alpha}$ values of $p(m; n\alpha)$, all approximately equal to $1/2\sqrt{n\alpha}$. The resulting estimate of the entropy $H_n(\alpha)$ of the discrete distribution $p(m; n\alpha)$ is

$$-2\sqrt{n\alpha}\,\frac{1}{2\sqrt{n\alpha}} \log \frac{1}{2\sqrt{n\alpha}} = \log 2\sqrt{n\alpha},$$

which diverges logarithmically as $n \to \infty$.

The appearance of this difficulty is not surprising since the entropy of a continuous random variable is defined as $-\int_{-\infty}^{\infty} p(\xi) \log p(\xi)\, d\xi$, where $p(\xi)$ is the probability density of the random variable, whereas the discrete random variable x_n used to define $H(\alpha)$ has no probability density, in spite of the fact that x_n converges in distribution to the continuous random variable x_∞. What has to be done is to replace the random variable x_n by a related continuous random variable x_n^* which converges to x_∞ in *density* as well as in distribution.[17] To achieve this, we define x_n^* as the continuous random variable with the probability density

$$p_n(\xi) = \sum_{m=0}^{\infty} \sqrt{n}\, p(m; n\alpha) h(\xi - y_{m,n}; 1/\sqrt{n}), \qquad (11)$$

where h is the step function defined by (1), and $y_{m,n} = m/\sqrt{n} - \sqrt{n}\alpha$ as usual. In other words, we replace the discrete random variable x_n by the continuous random variable x_n^*, where the "mass" formerly concentrated at the point $y_{m,n}$ is now uniformly distributed over the interval $(y_{m,n}, y_{m,n+1})$; of course, this requires adjustment of the step height, which accounts for the factor $\sqrt{n}$ in (11). Using a local central limit theorem,[18] we see that x_n^* converges in density to the Gaussian random variable x_∞, as $n \to \infty$. The entropy of x_n^* is defined in the usual way as

$$H_n^*(\alpha) = -\int_{-\infty}^{\infty} p_n(\xi) \log p_n(\xi)\, d\xi.$$

Substituting for $p_n(\xi)$ from (11), we find that

$$H_n^*(\alpha) = -\sum_{m=0}^{\infty} p(m; n\alpha) \log p(m; n\alpha) - \log \sqrt{n},$$

so that the logarithmic divergence of $H_n(\alpha)$ is cancelled out and the dependence of $H_\infty(\alpha)$ on α is restored. We should

[14] R. Fortet, "Random functions from a Poisson process," Proc. of the Second Berkeley Symp. on Math. Stat. and Prob. Theory, J. Neyman, Ed., University of California Press, Berkeley and Los Angeles, Calif. pp. 375–385; 1951.

[15] C. E. Shannon, "The Mathematical Theory of Communication," reprinted in the book of the same title by C. E. Shannon and W. Weaver, University of Illinois Press, Urbana, p. 54; 1949. Shannon's discussion of the difference between the entropy of a discrete random variable and that of a continuous random variable is highly relevant to the present analysis.

[16] Since $x_n(t)$ and $x_\infty(t)$ are stationary, we need not retain the argument t in discussing univariate quantities.

[17] As usual, we say that the random variable x_n converges in distribution as $n \to \infty$ to the random variable x_∞ if $F_n(\xi)$, the distribution function of x_n, converges to $F_\infty(\xi)$, the distribution function of x_∞, at every continuity point of the latter. If x_n and x_∞ have probability densities $p_n(\xi) = (d/d\xi)F_n(\xi)$ and $p_\infty(\xi) = (d/d\xi)F_\infty(\xi)$, respectively, we say that x_n converges in density to x_∞ if $p_n(\xi) \to p_\infty(\xi)$ almost everywhere. (Other definitions are possible.) Convergence in density implies convergence in distribution, but not conversely.

[18] B. V. Gnedenko and A. N. Kolmogorov, "Limit Distributions for Sums of Independent Random Variables," translated by K. L. Chung, Addison-Wesley Publishing Co., Cambridge, Mass., p. 233; 1954.

now have $H_n^*(\alpha) \to H_\infty(\alpha) = \log \sqrt{2\pi e\alpha}$ as $n \to \infty$. That this is the case is shown in Table I, where we give the results of numerical calculations of $H_n^*(\alpha)$ for the value $\alpha = 0.25$ used in constructing Figs. 1–3. (All logarithms are to the base e.) We see that as n increases, H_n^* (0.25) rapidly approaches the limiting value $H_\infty(0.25) = \log \sqrt{\pi e/2}$.[19] The univariate entropies of the processes $x_1(t)$, $x_4(t)$ and $x_{16}(t)$ represented in Figs. 1–3 are all rather close to the limiting value of 0.726, despite the fact that these cases correspond to low, medium and high density shot noise, respectively. However, the univariate entropy still seems to be a useful supplement to the power spectrum as an index of the structure of non-Gaussian random processes. (It will be recalled that all information about the relative phases of the harmonic components of a process are suppressed in its power spectrum, which limits the utility of the power spectrum as a means of characterizing non-Gaussian processes.)

[19] Note, however, the initial decrease of $H_n^*(0.25)$.

TABLE I[20]

n	$H_n^*(0.25)$
1	0.617
2	0.581
4	0.612
8	0.664
16	0.700
32	0.714
64	0.721
400[21]	0.722
∞	0.726

ACKNOWLEDGMENT

The author wishes to thank P. Elias for his very helpfu remarks concerning the entropy of $\{x_n(t)\}$.

[20] The numerical work was done by Miss P. A. Smith, using E. C. Molina's tables, "Poisson's Exponential Binomial Limit," D. Van Nostrand Co., Inc., New York, N. Y.; 1942.

[21] The value 400 corresponds to the largest value of $p(m; n\alpha$ available in Molina's tables.

On a Characterization of Processes for which Optimal Mean-Square Systems are of Specified Form*

A. V. BALAKRISHNAN†, MEMBER, IRE

Summary—**This paper presents the first results in an unconventional approach to the problem of mean-square optimization. Instead of obtaining a representation for the optimal operator for a process, we seek to characterize the class of processes for which the optimal operator is of specified form. If the processes are given, so that the multivariate characteristic functions are known, then our results can be used to tell whether it is possible for the optimal operator to have a specified form. The bulk of the paper pertains to the signal extraction problem where the signal and noise are independent and additive, and it is desired to estimate some function of the signal. Here, with a slight shift in viewpoint, we phrase the characterization problem in the following way: Given, for example, a noise process, determine the class of signal processes for which the optimal extraction system is of specified form. The case where the noise process is Gaussian comes in for special attention.**

I. INTRODUCTION

THE central problem in any mean-square optimization can be stated as follows: A noisy signal $Y(t)$, for t belonging to a parameter set π, where π may be discrete or continuous, is observed and it is desired to obtain the best estimate (in the mean-square sense) o the true signal $X(t_0)$ or some function of it, for example $Q[X(t_0)]$. Now, it is well-known that the optimal estimato is given by the conditional expectation

$$E\{Q[X(t_0)] \mid Y(t), t \in \prod\}.$$

This represents in general a nonlinear operation on $Y(t)$ and the problem has largely been studied under the re striction that the operations be linear. The reasons fo this are many, not the least of which is the difficulty i determining the optimal nonlinear operator in usabl form. Series expansions in canonical functions of som sort—polynomials for the most part—would appear to b unavoidable in order to represent the optimal operator i the general case [1–3]. As a possible alternate approach t this problem, we reverse the point of view and seek t characterize the class of processes for which a given systen or class of systems is optimal.[1] For example, we ma characterize the class of processes for which the optima system is specified in terms of a finite number of th

* Received by the PGIT, November 11, 1958; revised manuscript received, November 25, 1959.

† Space Technology Labs., Los Angeles, Calif.

[1] Our interest in the problem was rekindled during a conversatio with L. A. Zadeh. It is a pleasure to acknowledge our indebtednes to him.

anonical functions chosen for representation. (In this aper we consider only polynomials.) An advantage of he approach, apart from the intrinsic interest, is that the haracterization problem turns out to be a linear one in erms of the characteristic functions of the distributions nvolved.

Given a random process, what shall we mean by a haracterization of the process? If the parameter set is a nite set of discrete points, then, of course, the determi- ation of the multivariate density or distribution will haracterize the process completely. The characteristic unction, being the Fourier transform of the distribution, vill also be equally sufficient, since it determines the listribution uniquely. The basic relations derived in this aper are in terms of the characteristic functions. When ne parameter set is infinite, as in the case of a discrete r continuous parameter stochastic process, although the haracterization is, in principle, still possible in terms of an numeration of all the multivariate joint distributions, in ractice we have to make simplifying hypotheses allowing eduction in the variables to be specified. For example, or a Gaussian process the mean and correlation functions re sufficient. Alternately, we may assume that the process s Markoffian, and so on.

The characteristic functions, to be sure, lack physical ignificance and are not usually experimentally measured. On the other hand, the use of characteristic functions is nly an intermediate step in the problem, an analytical ool rather an end in itself.

One somewhat unexpected result of our theory is that nder certain conditions there will only be one process or which a given system is optimal. This opens the ossibility, for instance, of distinguishing between signals ased on the system for which it is optimal.

We begin in Section II with a set of necessary and ufficient conditions for multidimensional distributions n order that the optimal mean-square estimator be of pecified form. (We obtain this in terms of the character- stic functions which do not appear to be quite amenable o physical interpretations. Indeed, if physical intuition s inadequate in the "forward" problem, it is even more o in the "backward" problem considered here.) This ffords a means of checking whether or not a given process an lead to an optimal system of given form and has otential applications to all prediction and filtering roblems. In this paper, however, we confine our attention o the signal extraction problem where the signal and oise processes are additive and it is desired to obtain the ptimal mean-square estimate of some function of the ignal. Here we seek to characterize the class of signal rocesses for a given noise process and a given optimal xtraction filter. These process applications are given in ection III. We consider the zero-memory filters in some etail since they are not without importance in themselves nd indicate, moreover, the type of methods applicable nd the type of results to be expected in general. Our dis- ussion of the general case is confined to the optimal lters which are linear.

II. Characterization of Multivariate Distributions For Optimal Mean-Square Estimator of Prescribed Form

We begin by examining the conditions under which a multivariate distribution leads to an optimal estimator of prescribed form. By focusing attention on the corresponding characteristic functions, we obtain a set of usable necessary and sufficient conditions. We consider a finite number of (real) random variables $x_0, y_0, y_1, \cdots, y_n$, where x_0 is to be estimated in terms of $y_0, y_1, \cdots, y_n$. As is well-known (see [5], for example), the optimal mean-square estimate is then, of course, given by (the conditional expectation)

$$E[x_0 \mid y_0, \cdots y_n]. \tag{1}$$

In general, this is a Borel measurable function (in $n+1$ real variables). Let the characteristic function of the $n+1$ variables $y_0, \cdots, y_n$, be denoted $C_y(t_0, \cdots t_n)$, so that

$$C_y(t_0, \cdots t_n) = E\left[\exp \sum_0^n it_i y_i\right].$$

Then $C_y(\cdots)$ is a uniformly continuous function. Although for each particular result conditions can be relaxed, we shall now assume that all given random variables have finite moments of all orders. Then $C_y(\cdots)$ has derivatives of all orders and these are all again uniformly continuous. Let D_k denote the differential operator $\partial/\partial(it_k)$, so that

$$D_k C_y(t_0, \cdots t_n) = \frac{\partial}{\partial(it_k)} C_y(t_0, \cdots t_n).$$

(If we wish to use the formalism of the theory of linear operators on Banach spaces, we can take the space to be that of uniformly continuous functions on the Euclidean space E_{n+1} under the uniform norm, so that each D_k is a linear operator with dense domain.) In any case, any polynomial $P(D_0, \cdots D_n)$ is then also a well-defined linear operator, where $P(\cdots)$ is any polynomial in $(n+1)$ variates. The class of operator functions can be enhanced by taking limits of sequences of polynomial operators. This is as far as we shall go, since, for our purposes, the polynomials are the only ones on which we can have a direct and independent hold.

We begin with a theorem which is basic in our work. It is an extension of a result known in the special case where the estimator is linear, and the signal function estimated is also linear. (See [4], for example, where further reference may be found.)

Theorem 1: Let $P(\cdots)$ be a polynomial in $(n+1)$ variates (or, more generally, an entire function). Then, a necessary and sufficient condition for[2]

$$E[x_0^k \mid y_0, \cdots y_n] = P(y_0, \cdots y_n) \tag{2}$$

[2] All equalities of random variables are understood to be with probability one.

is that

$$\frac{\partial^k}{\partial(is_0)^k} C_{x_0,y}(s_0, t_0, \cdots t_n)\Big|_{s_0=0} = P(D_0, \cdots D_n)C_y(t_0, \cdots t_n) \tag{3}$$

where $C_{x_0,y}(s_0, t_0, \cdots, t_n)$ is the characteristic function of $x_0, y_0, y_1, \cdots y_n$.

Proof: We use a characteristic property of conditional expectations. (See [5] for example, p. 22). This is to the effect that if $g(\cdots)$ is any Borel measurable function in $(m + 1)$ variables and

$$E[|\, x_0^k g(y_0, \cdots y_n)\,|] < \infty,$$

then

$$E[x_0^k g(y_0, \cdots y_n)] = E\{g(y_0, \cdots y_n)E[x_0^k \mid g(y_0, \cdots y_n)]\}. \tag{4a}$$

Specializing this to the case where

$$g(y_0, \cdots y_n) = \exp i \sum_0^n t_k y_k,$$

we have

$$E[x_0^k \exp i \sum t_k t_k] = E\left\{\left(\exp i \sum_0^n y_k t_k\right) E[x_0^k \mid y_0, \cdots y_n]\right\}. \tag{4b}$$

Since kth moments are finite, we can differentiate inside the expected value integral on the left in (4b), so that

$$E\left[x_0^k \exp i \sum_0^n t_k y_k\right] = \frac{\partial^k}{\partial(is_0)^k} C_{x_0,y}(s_0, t_0, \cdots t_n)\Big|_{s_0=0}. \tag{5}$$

Next, let (2) hold with $P(\cdots)$ a polynomial. Again, if the degree of $P(\cdots)$ is equal to n and all nth moments are finite (as we are assuming), we can differentiate inside the expected value integral on the right side of (4b) so that we have

$$E[P(y_0, \cdots y_n) \exp i \sum t_k y_k] = P(D_0, \cdots D_n)C_y(t_0, \cdots t_n). \tag{6}$$

Combining (4b), (5) and (6), we obtain (3). Conversely, suppose (3) is true. Since the necessary moments are assumed finite, we obtain (6) and (5). Substituting in (4b), we have

$$E\left\{E[x_0^k \mid y_0, \cdots y_n] \exp i \sum_0^n t_k y_k\right\} = E\left[P(y_0, \cdots y_n) \exp i \sum_0^n t_k y_k\right].$$

Since this holds for every choice of $\{t_k\}$, (2) follows from the uniqueness theorem for Fourier-Stieltjes transforms. Extension to entire functions can be made by taking limits of polynomials.

We next consider the special case where

$$y_k = x_k + N_k$$

and x_k, N_j are stochastically independent. We have the Theorem 2.

Theorem 2: Let $Q(\cdot)$ be a polynomial in one variabl and let $P(\cdots)$ be a polynomial in $(n + 1)$ variates. Ther a necessary and sufficient condition that

$$E[Q(x_0) \mid y_0, y_1, \cdots y_n] = P(y_0, y_1, \cdots y_n) \tag{7}$$

is that

$$C_N(t_0, \cdots t_n)Q(D_0)C_x(t_0, \cdots t_n) = P(D_0, \cdots D_n)C_x(t_0, \cdots t_n)C_N(t_0, \cdots t_n). \tag{8}$$

Proof: We first note that

$$C_{x_0,y}(s_0, t_0, \cdots t_n) = C_N(t_0, \cdots t_n)C_x(s_0, t_0, \cdots t_n)$$

so that

$$\frac{\partial^k}{\partial(is_0)^k} C_{x_0,y}(s_0, t_0, \cdots t_n)\Big|_{s_0=0} = C_N(t_0, \cdots t_n) \cdot D_0^k C_x(t_0, \cdots t_n)$$

Let $Q(x) = \sum a_j x^j$, then

$$E[Q(x_0) \mid y_0, \cdots y_n] = \sum a_j E[x_0^j \mid y_0, \cdots y_n]$$

so that, using (5) term by term, and substituting in (3) we have

$$C_N(t_0, \cdots t_n)Q(D_0)C_x(t_0, \cdots t_n) = P(D_0, \cdots D_n)C_n(t_0, \cdots t_n)$$

But since x_k, N_j are independent,

$$C_y(t_0, \cdots t_n) = C_x(t_0, \cdots t_n)C_N(t_0, \cdots t_n),$$

so that (8) follows. Extensions to the case where $Q(\cdot)$ $P(\cdots)$ are entire functions can be made by taking limits

Theorems 1 and 2 have assumed the optical estimato to be a polynomial or limits of such because then th corresponding linear operator on the characteristic func tions can be defined independently without bringing i the distributions. This can, of course, be extended t estimators which are not necessarily entire functions pro vided there is still a sequence of polynomials convergin to the optimal estimator in the mean of order two. Th latter would be the case, for instance, if the polynomial are complete in the L_2 space induced by the joint distri bution of the $\{y_k\}$. However, since this is not true withou additional assumptions and since, in this paper, w intend only to indicate the possibilities of our approach we confine ourselves more or less exclusively to poly nomials in the next Section.

III. Process Applications

Theorem 1 would appear to be of potential applicatio to any prediction or extraction problem. In the firs place, if the problem can be reduced to one involvin

finite-dimensional distributions[3] and it is desired to check whether it is possible for the system to be of specified form, then Theorem 1 can be used directly to provide the answer. However, shifting to a larger view, we now consider the class of processes for which a given system (or class of system) is optimal. Moreover, we restrict ourselves in what follows to the signal extraction problem where the signal and noise processes are independent and additive, and some function of the signal is to be estimated. Thus, let $X(t)$ represent the signal and $N(t)$ the noise which is independent of the signal. Let π be the parameter set and

$$Y(t) = X(t) + N(t), \qquad t \in \pi.$$

Let $Q(\cdot)$ be a (Borel) function. We now phrase the characterization problem in the following way: For a given noise process $N(t)$, characterize the class of signal processes $X(t)$ for which the optimal estimator

$$E\{Q[X(t)] \mid Y(t),\ t \in \pi\}$$

is specified in form. In this paper, we consider only the case where $Q(\cdot)$ is a polynomial, and the sample-point set π is (or can more or less be reduced to be) finite, and moreover, the optimal estimator is also a polynomial. It is then clear from (8) that this would amount to solving a partial differential equation for $C_x\ (t_0, \cdots t_n)$ for given $C_N\ (t_0, \cdots t_n)$. The important feature is that the differential equation is linear.

Zero-Memory Filters

We begin with optimal zero-memory filters. These are the simplest to consider since they involved only one sample point and, hence, only first order distributions. Nevertheless, they illustrate the type of results to be expected in the general case. Moreover, they constitute the essential part of an important class of nonlinear filters which consist of a zero-memory device sandwiched between two linear filters [3].

Let $Q(\cdot)$, $P(\cdot)$ be polynomials. Then, for any t_0, we wish to consider the class of processes for which

$$E\{Q[X(t_0)] \mid Y(t_0)\} = P[Y(t_0)]. \tag{9}$$

Specializing Theorem 2, this means that we must have

$$C_N(t)Q(D)C_x(t) = P(D)[C_x(t)C_N(t)] \tag{10}$$

where $C_x(t)$, $C_N(t)$ are the characteristic functions of $X(t_0)$ and $N(t_0)$, respectively, and $D = 1/i\ d/dt$. For given $C_N(t)$, (10) is then an ordinary differential equation for $C_x(t)$, and can be solved subject to the condition that the solution be a characteristic function. Before discussing (10) in full generality, we shall first explore some simple cases. First, let

$$P(D) = a\ D \tag{11}$$

so that $C_x(t)$ must satisfy the differential equation

$$Q(D)C_x(t) - a\ DC_x(t) = a\left[\frac{DC_N(t)}{C_N(t)}\right]C_x(t)$$

$$= a[D\chi_N(t)]C_x(t), \tag{12}$$

$\chi_N(t)$ being the logarithm of $C_N(t)$. Further specializing $Q(D)$ to be linear, we have (see [4]) Theorem 3.

Theorem 3: Suppose

$$E[X(t_0) \mid Y(t_0)] = aY(t_0). \tag{13}$$

Then, for some $\delta > 0$,

$$C_x(t) = [C_N(t)]^\alpha, \qquad 0 \leq \alpha = \frac{a}{1-a},$$

$$\text{for} \quad 0 \leq |\ t\ | \leq \delta \tag{14}$$

where the determination of $[C_N(t)]^\alpha$ is such that it is positive when $C_N(t)$ is positive. Conversely, if (13) is satisfied for all (t), then (14) again holds.

Proof: The proof is immediate, since (12) simplifies to

$$C_N(t)(1-a)\ DC_x(t) = aC_N(t)\ DC_x(t)$$

so that

$$\frac{C_x'(t)}{C_x(t)} = \frac{a}{1-a}\frac{C_N'(t)}{C_N(t)} \tag{15}$$

where the primes denote differentiation with respect to t. For small enough δ, $\log C_N(t)$ can be defined so that it is real when $C_N(t)$ is positive. Hence, we obtain from (15)

$$\log C_x(t) = \left(\frac{a}{1-a}\right)\log C_N(t),$$

from which (14) follows, thus proving the necessity. The converse is clear, since we have only to retrace the steps.

For the sake of completeness and at the same time to note some of the characteristic features of the problem, we shall now detail a few examples. It is hardly necessary to add that (13) is equivalent to saying that the optimal filter is linear. Hence, (14) characterizes the processes for which the optimal zero-memory filter is linear. In practice, it is certainly no restriction to assume that $C_N(t)$ has a MacLaurin expansion in a nonzero interval about the origin. If (14) holds, this then implies that $C_x(t)$ has also a similar expansion and, thus by a known result due to Marcinkiewicz (see [6], p. 212), it follows that $C_x(t)$ for other values of t are determined uniquely. It follows by inspection that if the noise is Gaussian or Poisson, then the optimal filter is linear if, and only if, the signal is also Gaussian or Poisson, respectively. As an example of a distinctly non-Gaussian[4] process, we mention the case

[3] The initial Laguerre transformation in [2] has this effect, for example.

[4] More generally, a class of distributions known as "infinitely divisible" distributions (see [6]) are left invariant, that is to say if one of the processes has an infinitely divisible distribution, so must the other for the optimal filter to be linear. However, the only practical example of the infinitely divisible distributions are the Gaussian and Poisson, and the Γ-type already cited.

where the noise is, for example, Γ type (this terminology is found in [6], p. 215), as when

$$C_N(t) = [1 - cit]^{-\gamma}, \gamma, \qquad c > 0.$$

Then the optimal filter is linear if and only if $C_x(t)$ is also of the Γ type with

$$C_x(t) = [1 - cit]^{-a\gamma}.$$

We note that the corresponding densities vanish on the negative real axis.

Let us note that in order for $C_x(t)$ given by (14) to be a characteristic function, it is necessary that $0 \leq a \leq 1$. However, for any given "a" in this range, there need not be a characteristic function. As an example, we have only to take

$$C_N(t) = [p + qe^{it}]^n, \quad \text{with} \quad p + q = 1,$$

$$0 \leq p, \qquad q \leq 1,$$

and in this case it is clear that $n\alpha$ would have to be an integer in order for $C_x(t)$ to be a characteristic function. On the other hand, there is at most one characteristic function solution (or signal process) since the differential equation (15) is of the first order.

It should be pointed out that saying that the optimal filter is linear implies that there is a lower bound to the error in filtering (this error is readily seen to be

$$\frac{[\text{variance of } X(t_0)][\text{variance of } N(t_0)]}{\text{variance of } X(t_0) + \text{variance of } N(t_0)}\Bigg).$$

We have noted the features of the problem when the optimal filter is required to be a linear carry-over, *mutatis mutandis*, to the nonlinear situation. Unfortunately, it is not possible to state general results covering all nonlinearities postulated and this is a phenomenon familiar in the theory of nonlinear systems. We shall therefore illustrate the type of results to be expected with examples.

We shall first consider (12) to show that it is not always possible to find characteristic function solutions for any choice of type of optimal filter. Thus, let the filter be nonlinear so that the polynomial $Q(D)$ in (12) is of an order higher than one. Let us take the noise to be Gaussian, so that

$$C_N(t) = \exp - \frac{\lambda^2 t^2}{2}.$$

Then we have, from (12),

$$Q(D)C_x(t) - a\, DC_x(t) = a\lambda^2(it)C_x(t). \tag{16}$$

This equation can be solved explicitly, since the left-hand side has constant coefficients. Thus, let

$$Q(D) = \sum_1^n b_k\, D^k.$$

Then, letting $f(x)$ be such that

$$\int_{-\infty}^{\infty} e^{itx} f(x)\, dx = C_x(t),$$

we have, by taking (inverse) Fourier transforms in (16) that

$$\left[\sum_1^n b_k x^i - ax\right] f(x) = -a\lambda^2 \frac{df(x)}{dx},$$

and hence

$$f(x) = C \exp - \int_0^x \frac{\left[\sum_1^n b_k y^k - ay\right]}{a\lambda^2}\, dy. \tag{17}$$

It follows from this at once that n [$n \geq 2$] has to be odd and further that (b_n/a) must be positive, and moreover that this is enough. The constant C in (17) has evidently to be chosen so that $f(x)$ integrates to unity. Comparison of the density function (17) with the Gaussian shows that it is steeper than the Gaussian, going to zero more rapidly. However, physical intuition appears to be inadequate to pin down the precise nature of the distribution of $X(t_0)$. It is also readily shown from (16) that the solution obtained is the only possible distribution (with all moments finite).

As another example of (12), we may take a discrete distribution and let $N(t_0)$ be Poisson so that

$$C_N(t) = \exp \lambda[e^{it} - 1].$$

Then let

$$C_x(t) = \sum_0^\infty \beta_m e^{itm}.$$

If we substitute this into (12), we obtain

$$\left[\sum_1^n b_k m^k - am\right]\beta_m = [a\lambda]\beta_{m-1}, \quad m = 1, 2, \cdots. \tag{18}$$

This equation determined $\{\beta_m\}$ and it is readily seen that we require

$$\left\{\frac{\sum_1^n b_k m^k - am}{a\lambda}\right\} > 0$$

for every positive integer and, moreover, this condition is sufficient.

The last two examples can be extended to functions $Q(\cdot)$ which are not polynomials. Thus, in (17) we have only to replace $\sum a_k\, y^k$ by $Q(y)$, and similarly in (18), and the corresponding extended conditions on the function $Q(t)$ for (17) to yield a probability density and for (18) to yield a (discrete) distribution can be readily determined.

Let us now return to the general case where $Q(\cdot)$ and $P(\cdot)$ are arbitrary polynomials. First, we note that if we set

$$Q(D) = \sum_1^m b_k\, D^k$$

and

$$P(D) = \sum_1^n a_k\, D^k,$$

We can use the Leibnitz rule for differentiating a product and rewrite (10) as

$$D^N\left[\sum_0^m b_i D^i C_x\right] = \sum_{i=0}^{n}\left[\sum_{k=i}^{n} a_k C_i^k D^{k-i} C_N\right] D^i C_x \qquad (19)$$

where

$$C_i^k = \frac{k!}{j!\,(k-j)!}.$$

This is a homogeneous ordinary linear differential equation with variable coefficients and the coefficients are all bounded functions. In general, (19) can have, of course, n linearly independent solutions that are characteristic functions and any convex linear extension of these will again be a characteristic function. Thus, if (19) has two linearly independent solutions that are characteristic functions, it will then have infinitely many.

We shall now illustrate the use of a technique that leads to solutions of a class of characterization problems where we are interested in invariance—that is, in the cases where both signal and noise distributions belong to the same class, Gaussian for instance. For this, we use (10) directly and appropriate polynomial expansions of the distribution involved.

Suppose, then, that the noise process is Gaussian with

$$C_N(t) = \exp -\frac{\lambda_1^2 t^2}{2}$$

and we are interested in the conditions on the polynomial $Q(\cdot)$ and $P(\cdot)$ that will ensure that the signal $X(t)$ is also Gaussian. Let us, thus, assume that

$$C_x(t) = \exp -\frac{\lambda_2^2 t^2}{2}$$

and look at the consequences. First, let us note that

$$D^k C_N(t) = (-i)^n \frac{d}{dt^n} \exp -\frac{\lambda_1^2 t^2}{2}$$

$$= [(i\lambda_1)^n H_n(t\lambda_1)] \exp -\frac{\lambda_1^2 t^2}{2}$$

where $H_n(\cdot)$ is the nth Hermite polynomial

$$H_n(x) = e^{x^2/2}(-1)^n \frac{d}{dx^n} \exp -x^2/2$$

using the notation

$$\lambda^2 = \lambda_1^2 + \lambda_2^2.$$

We then have from (10)

$$C_N Q(D) C_x = \left[\sum_1^m b_k (i\lambda_2)^k H_k(\lambda_2 t)\right]\left[\exp -\frac{\lambda^2 t^2}{2}\right]$$

and

$$P(D)[C_N C_x] = \left[\sum_1^n a_k (i\lambda)^k H_k(\lambda t)\right]\left[\exp -\frac{\lambda^2 t^2}{2}\right].$$

Hence, we must have, for (10) to hold,

$$\sum_1^n a_k (i\lambda)^k H_k(\lambda t) = \sum_1^m b_k (i\lambda_2)^k H_k(\lambda_2 t). \qquad (20)$$

This is an identity in t so that we can equate like powers of t. Clearly, there is no harm in taking $m = n$, since we can consider the necessary coefficients to be zero. Since the odd order Hermite polynomials are odd, and the even order polynomials even, we can separate them in (20). To be specific, suppose $n = 3$. Then we have

$$\begin{aligned} a_3\lambda^6 &= b_3\lambda_2^6, \\ a_1\lambda^2 + 3a_3\lambda^4 &= b_1\lambda_2^2 + 3b_3\lambda_2^4, \\ a_2\lambda^4 &= b_2\lambda_2^4, \end{aligned} \qquad (21)$$

and

$$\lambda^2 a_2 = \lambda_2^2 b_2.$$

These conditions can be satisfied only if $a_2 = b_2 = 0$, omitting the trivial case of zero variance. In other words, for $n = 3$, $P(\cdot)$ and $Q(\cdot)$ must be odd functions. For $n = 5$ and higher, this, of course, need not be. We may formalize this result as Theorem 4.

Theorem 4: Let $X(t_0)$ and $N(t_0)$ have zero means and let their variances be λ_2^2 and λ_1^2, respectively. Let $N(t_0)$ be Gaussian. Then

$$E\{Q[X(t_0)] \mid Y(t_0)\} = P[Y(t_0)]$$

where $P(\cdot)$ and $Q(\cdot)$ satisfy (20) if, and only if, $X(t_0)$ is also Gaussian.

Proof: Since we have already given the arguments for the "if" part, we need only prove necessity. Suppose then that (20) is satisfied. Then one solution of (10) is clearly gotten by taking

$$C_x(t) = \exp -\frac{\lambda_1^2 t^2}{2}.$$

We have now to show that this is the only solution. This is given in Appendix I.

It is, of course, possible to resolve (20) into (21) for arbitrary "n," but since this involves too much notation, we have refrained from doing so.

As another example, we shall consider the case where the noise has a Γ-type distribution. Thus, let

$$C_N(t) = [1 - cit]^{-\gamma_1}; \qquad (22)$$

then note that

$$D^k[C_N(t)] = \frac{c^k \gamma_1(\gamma_1 + 1) \cdots (\gamma_1 + k - 1)}{[1 - cit]^{+\gamma_1 + k}}. \qquad (23)$$

Proceeding as in the previous example, we can state Theorem 5.

Theorem 5: Let $N(t_0)$ have Γ-type distribution so that $C_N(t)$ is given by (22). Then a necessary and sufficient condition for

$$E[X(t_0)^n \mid Y(t_0)] = \alpha[Y(t_0)]^n, \quad n \text{ positive integer}, \qquad (24)$$

where $0 < \alpha < 1$ is that $X(t_0)$ also have a distribution of the Γ type.

Proof:

Sufficiency: Let

$$C_x(t) = [1 - ict]^{-\gamma_2}.$$

Then we shall show that for suitably chosen γ_2 (24) holds. For this, note that (10) in this case, using (22), can be written

$$c^n\gamma_1(\gamma_1 + 1) \cdots (\gamma_1 + n - 1) = \alpha c^n\gamma(\gamma + 1) \cdots (\gamma + n - 1) \quad (25)$$

where

$$\gamma = \gamma_1 + \gamma_2.$$

Since $o < \alpha < 1$, it is clear that (25) can be solved to yield a positive γ, $\gamma > \gamma_1$, and hence, $\gamma_2 = \gamma - \gamma_1$ will serve as the exponent.

Necessity: Suppose that (24) holds. Then we know that by Theorem 2 we must have

$$C_N\, D^n C_x = \alpha\, D^n[C_N C_x]. \quad (26)$$

Now, into (26), let us put

$$C_x(t) = [1 - itc]^{-\gamma_2}.$$

Then, in order for this $C_x(t)$ to be a solution of (26), it is enough if (25) is true. Now (25) considered as an equation for γ has exactly n roots, each root γ yielding $(\gamma - \gamma_1)$ as a possible value for γ_2. However, the corresponding solution will be a characteristic function if and only if γ_2 is positive. For $0 < \alpha < 1$, as given, (25) has a positive root γ, $\gamma > \gamma_1$. Moreover, this is the only such root. For, upon differentiating (25) with respect to γ, we note that the derivative is positive for all positive γ, and hence, (25) cannot have more than one positive root. This concludes the proof of the theorem.

Let it be noted that we have proved slightly more than the theorem states. Thus, the exponent γ_2 is to be calculated from (25).

Filters with Memory

In the more general case, which we now consider, the optimal filters will have memory. Now it is clear that our methods can yield information on the joint distribution of the processes involved for as many sample points as are used in the optimization, excluding trivial cases. The theory is naturally richer since there are more degrees of freedom. Here we discuss primarily the characterization problem for processes for which the optimal extraction filter is linear. The partial differential equations that arise are then of the first order and our result pertains to their solution. We again assume that the distributions have all finite moments of as high an order as required.

Theorem 6: Let

$$E[X(t_0) \mid Y(t_0), Y(t_1), \cdots Y(t_n)] = (1 - a_0)Y(t_0) + \sum_1^n a_k Y(t_k) \quad (27)$$

where not all $\{a_i\}$ are zero, for example, $a_1 \neq 0$. Then for some $\delta > 0$ and all $\{t_i\}$, such that

$$|\, t\,| = \sqrt{t_o^2 + t_1^2 + \cdots t_n^2} < \delta,$$

the characteristic function of $X(t_0)$, $X(t_1)$, $\cdots X(t_n)$ must be of the form

$$\log C_x(t_0, t_1, \cdots t_n) = R(t_1, c_1, \cdots c_n) + g(c_1, c_2, \cdots c_n) \quad (28$$

where

$$\begin{aligned} c_1 &= a_0 t_1 + a_1 t_0, \\ c_2 &= a_2 t_1 - a_1 t_2, \\ &\vdots \\ c_n &= a_n t_1 - a_1 t_n, \end{aligned} \quad (29$$

$$R(t_1, c_1, \cdots c_n) = -\frac{1}{a_1}\int_0^{t_1} Q\left[\frac{a_0 t_1 - c_1}{a_1}, t_1, \frac{a_2 t_1 - c_2}{a_1}, \cdots \frac{a_n t_1 - c_n}{a_1}\right] dt_1$$

$$Q(t_0, t_1, \cdots t_n) = \left[(1 - a_0)\frac{\partial}{\partial t_0} - \sum_1^n a_k \frac{\partial}{\partial t_k}\right] \log C_N(t_0, t_1, \cdots t_n$$

and $g(\cdots)$ is an arbitrary function. Conversely, if (28 satisfies for every $\{t_i\}$, then (27) holds.

Proof:

Necessity: By Theorem 2, for (27) to hold, we must hav

$$C_N(t_0, t_1, \cdots t_n)\left[a_0 \frac{\partial}{\partial t_0} - a_1 \frac{\partial}{\partial t_1} \cdots a_n \frac{\partial}{\partial t_n}\right] C_x(t_0, t_1, \cdots t_n) = C_x(t_0, \cdots t_n)\left[(1 - a_0)\frac{\partial}{\partial t_0} + a_1 \frac{\partial}{\partial t_1} + \cdots + a_n \frac{\partial}{\partial t_n}\right] C_N(t_0, t_1, \cdots t_n). \quad (30$$

As before, we may consider this as a (partial) differ ential equation for $C_x(t_0, t_1, \cdots t_n)$ with $C_N(t_0, t_1, \cdots t_n$ given. The equation is of the first order and can be solve by standard methods. We include a few steps in th solution. First, we can choose δ so that for $|t| < \delta$, both characteristic functions are never zero or negative, so tha their logarithms can be defined. Next, letting

$$Q(t_0, t_1, \cdots t_n) = \left[(1 - a_0)\frac{\partial}{\partial t_0} + a_1 \frac{\partial}{\partial t_1} \cdots + a_n \frac{\partial}{\partial t_n}\right] \text{Log } C_N(t_0, t_1, \cdots t_n)$$

and

$$F = \log C_x(t_0, t_1, \cdots t_n),$$

we have

$$\frac{dt_0}{a_0} = \frac{dt_1}{-a_1} = \cdots \frac{dt_n}{-a_n} = \frac{dF}{Q}.$$

Defining $\{c_i\}$ as in (29), we can express $t_0, t_2, t_3, \cdots t_n$ in terms of $\{c_i\}$ and t_1, since a_1 is assumed nonzero, as follows:

$$\begin{aligned} t_0 &= \frac{a_0 t_1 - c_1}{-a_1}, \\ t_2 &= \frac{a_2 t_1 - c_2}{a_1}, \\ &\vdots \\ t_n &= \frac{a_n t_1 - c_n}{a_1}. \end{aligned} \tag{31}$$

Substituting these in $Q(t_0, t_1, \cdots t_n)$ and setting

$$R(t_1, c_1, c_2, \cdots c_n) = -\frac{1}{a_1}\int_0^{t_1} Q\left[\frac{a_0 t_1 - c_1}{-a_1}, t_1, \cdots, \frac{a_n t_1 - c_n}{a_1}\right] dt_1,$$

we obtain the general solution of (30) as

$$R(t_1, c_1, \cdots c_n) + g(c_1, c_2, \cdots c_n).$$

Moreover, since $a_1 \neq 0$, any solution of (30) must be so expressible. Hence (28) follows. Here $g(c_1, c_2, \cdots c_n)$ is the general solution of the homogeneous equation and has as many derivatives as $C_x(t_0, t_1, \cdots t_n)$ has except where the latter is zero.

Conversely, suppose (28) is true. Then by direct differentation (30) follows and by Theorem 2 (27) also follows.

Suppose now that, for example, the noise process distribution is specified. Then for arbitrarily given $\{a_i\}$ there is an $X(t)$ process for which (27) holds, only if there is a characteristic function of the form given by (28). Suppose, in particular, that the noise is Gaussian. Then, we know that if there is a solution at all, there is always a Gaussian solution which has the same first and second moments as the given solution. On the other hand, in contrast to the one-dimensional problem, there may be more than one solution even in this case. We illustrate this with the following example. We consider an optimal extraction filter with two sample points. Thus, let

$$C_N(t_0, t_1) = \exp - \tfrac{1}{2}(t_0^2 + t_1^2)$$

and

$$G(t_0, t_1) = \exp - \tfrac{1}{2}(t_0^2 + t_1^2 + t_0 t_1).$$

Then, $G(t_0, t_1)$ satisfies (30), considered as an equation for $C_x(t_0, t_1)$ with $C_N(t_1, t_1)$ given as above, if we set

$$a_0 = 8/15, \qquad a_1 = 2/15$$

and the rest of the $\{a_i\}$ zero.

Next let

$$C(t_0, t_1) = G(t_0, t_1)\left[1 + \frac{\alpha_4}{4!}\left(\frac{2t_0 + 8t_1}{15}\right)^4\right].$$

Then $C(t_0, t_1)$ also satisfies (30) with the same choice of a_0 and a_1, for every value of α_4. We shall now show that $C(t_0, t_1)$ is a characteristic function with a suitable value for α_4. For this, let

$$T = (\text{the column vector}) \begin{vmatrix} t_0 \\ \sigma \end{vmatrix},$$

$$Q(t_0, \sigma) = T'AT$$

(T' being the transpose of T), and

$$A = \begin{vmatrix} 1 & -2/8 \\ 0 & 15/8 \end{vmatrix} \begin{vmatrix} 1 & 1/2 \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} \begin{vmatrix} 1 & 0 \\ -2/8 & 15/8 \end{vmatrix} = \frac{1}{64}\begin{vmatrix} 52 & 30 \\ 30 & 225 \end{vmatrix}.$$

Then

$$A^{-1} = \frac{4}{675}\begin{vmatrix} 225 & -30 \\ -30 & 52 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}.$$

Now let

$$C_0(t_0, \sigma) = \left[1 + \frac{(a_{22})^{-2}}{4!}\sigma^4\right] \exp - \tfrac{1}{2}Q(t_0, \sigma)$$

so that its inverse Fourier transform is

$$\frac{1}{(2\pi)^2}\iint C_0(t, \sigma) \exp - i(xt + \sigma y)\, dt\, d\sigma = G_0(x, y)\left\{1 + \frac{1}{4!}H_4\left[\sqrt{a_{22}}\left(y + \frac{a_{12}}{a_{11}}x\right)\right]\right\}$$

where $H_4(\cdot)$ is the fourth Hermite polynomial as defined before, and $G_0(x, y)$ is the Gaussian density corresponding to $\exp - \frac{1}{2}Q(t_0, \sigma)$. Since

$$\left[1 + \frac{\alpha_4}{4!}H_4(x)\right] \geq 0, \qquad -\infty < x < \infty,$$

provided $0 \leq \alpha_4 \leq 4$, it follows that $C_0(t_0, \sigma)$ is the Fourier transform of a nonnegative function, and since $C_0(0, 0) = 1$, it is a characteristic function. But

$$C(t_0, t_1) = C_0\left[t_0, \frac{2t_0 + 8t_1}{15}\right]$$

so that $C(t_0, t_1)$ is a characteristic function also, as required.

However, under certain conditions, the linearity of the optimal filter does imply that $X(t)$ is Gaussian, also, as will be shown in Theorem 7.

Theorem 7: Let $N(t)$, $t \in \pi$ be Gaussian and let $E(N(t)^2] \neq 0$. Then a necessary and sufficient condition that $X(t)$, $t \in \pi$ be Gaussian also is that

$$\hat{E}[X(t) \mid Y(t), Y(t - T_1), \cdots Y(t - T_n)] = \hat{E}[X(t) \mid Y(t), Y(t - T_1), \cdots Y(t - T_n)] \tag{32}$$

for every $t \in \pi$ and for every choice of distinct $\{T_i\}$ and n, $(t - T_i) \in \pi$, where $\hat{E}$ denotes the conditional expectation assuming Gaussian distributions have the same means and variances as the given processes, and is the optimal linear filter.

Proof: The sufficiency part being well-known, we need only prove necessity. We prove this by induction. Suppose, then, we have proved that all n-dimensional distributions of the $X(t)$ process are Gaussian. We shall first show that this implies that all $(n + 1)$ dimensional distributions be Gaussian also. We construct a Gaussian process with the same first and second moments as the given $X(t)$ process. Let us pick any set of points $t, t - T_1, \cdots t - T_n$ from π. Let the corresponding characteristic function of $X(t)$, $X(t - T_1), \cdots X(t - T_n)$ be denoted $C_x(t_0, t_1, \cdots t_n)$. Let $G(t_0, t_1, \cdots t_n)$ be the constructed Gaussian characteristic function corresponding to $C_x(t_0, t_1, \cdots t_n)$. Let

$$E[X(t) \mid Y(t), Y(t - T_1), \cdots Y(t - T_n)]$$
$$= (1 - a_0)Y(t) + a_1 Y(t - T_1) + \cdots + a_n Y(t - T_n). \quad (33)$$

Since $N(t)$ is a Gaussian process, $G(t_0, t_1, \cdots t_n)$ is, of course, a particular solution of (30) which $C_x(t_0, \cdots t_n)$ must satisfy by Theorem 2. Now, not all $\{a_i\}$ in (33) can be zero. For in that case, since

$$E[N(t) \mid Y(t), \cdots Y(t - T_n)]$$
$$= Y(t) - E[X(t) \mid Y(t), Y(t - T_1), \cdots Y(t - T_n)]$$
$$= A_0 Y(t) - \sum_1^n a_k Y(t - T_k),$$

this conditional expectation would be zero also. This implies that $E[N(t)^2] = 0$, contrary to the assumption that $E[N(t)^2] \neq 0$. Hence, from Theorem 5, assuming, for example that for $0 \leq t \leq \delta$

$$C_x(t_0, t_1, \cdots t_n) = G(t_0, t_1, \cdots t_n)h(t_0, t_1, \cdots t_n), \quad (34)$$

where $h(t_0, \cdots t_n)$ is a solution of the homogeneous equation

$$\left[(1 - a_0)\frac{\partial}{\partial t_0} + a_1 \frac{\partial}{\partial t_1} + \cdots + a_n \frac{\partial}{\partial t_n}\right] h(t_0, \cdots t_n) = 0.$$

Actually, a little consideration of (30) shows that (34) holds for all $\{t_i\}$ since a particular solution valid for all $\{t_i\}$ is known. On the other hand, if, for example, a_1 is assumed to be nonzero, $h(t_0, \cdots t_n)$ must be of the form

$$h(t_0, \cdots t_n) = g(c_1, c_2, \cdots c_n)$$

where the $\{c_i\}$ are given again by (29). If in (34) we set $t_1 = 0$ we have

$$C_x(t_0, 0, t_2, \cdots t_n)$$
$$= G(t_0, 0, t_2, \cdots t_n)g(a_1 t_0, -a_1 t_2, -a_1 t_3, \cdots -a_1 t_n).$$

But by the induction hypothesis, the left-hand side is Gaussian and hence equal to the first factor on the right. Hence

$$G(a_1 t_0, -a_1 t_2, -a_1 t_3, \cdots, -a_1 t_n) \equiv 1$$

or

$$h(t_0, t_1, \cdots t_n) = 1$$

also. Hence $C_x(t_0, t_1, \cdots t_n)$ is Gaussian. To complete th induction, we have only to show that first-order distri butions are Gaussian, and this readily follows fro Theorem 3.

An obvious implication of Theorem 7 is, of course, tha in the multidimensional case we must require that a finite sample point optimal filters be linear, and, of course in the counter example given prior to the theorem, thi is not true. We may note that Theorem 7 also holds if, fo instance, we place "Gaussian" therein by "Poisson. We have, of course, made no stationarity assumption.

If the processes are strictly stationary, we can weake the conditions of Theorem 7, as will be shown in Theorem 8

Theorem 8: Let $N(t)$, $X(t)$, $-\infty < t < \infty$ be strictl stationary stochastic processes. Let $N(t)$ be Gaussian wit nondegenerate second-order densities.[5] Let $X(t)$ be Markoff process. Then a necessary and sufficient conditio for

$$E[X(t) \mid Y(t), Y(t - T)] = \hat{E}[X(t) \mid Y(t), Y(t - T)] \quad (35$$

for every $T > 0$ is that $X(t)$ be Gaussian.

Proof: Sufficiency being again trivial, we need onl prove necessity. Let $C_x(t_0, t_1)$ be the characteristic func tion of $X(t), X(t - T)$ and $C_N(t_0, t_1)$ that of $N(t), N(t - T$ Let

$$E[X(t) \mid Y(t), Y(t - T)]$$
$$= (1 - a_0)Y(t) + a_1 Y(t - T). \quad (36$$

Now a_0 and a_1 cannot both be zero since $E[N(t)^2 \neq 0$ b hypothesis. Let $G(t_0, t_1)$ be the Gaussian characteristi function having the same first and second moments a $C_x(t_0, t_1)$. As in Theorem 7, we know that $C_x(t_0, t_1)$ mus be of the form

$$C_x(t_0, t_1) = G(t_0, t_1)h(a_0 t_1 + a_1 t_0).$$

However, by assumed stationarity,

$$C_x(0, t) = G(0, t)h(a_0 t)$$
$$= C_x(t, 0) = G(t, 0)h(a_1 t).$$

Hence

$$h(a_1 t) = h(a_0 t) \quad (37$$

for all t. If either a_0 or $a_1 = 0$, this would imply tha $h(t) = 1$ or that $C_x(t_0, t_1)$ is Gaussian. Since both canno be zero, we need only consider the case where neither a nor a_0 is zero. Suppose now that $h(t)$ is not a constan

[5] By this we mean that the matrix of second moments is *positiv* definite.

hen (37) clearly implies that $a_1 = a_0$. However, this leads o a contradiction, for since

$$a_o\lambda_{00} - a_1\lambda_{01} = M_{00} + E[X(t)]E[N(t)]$$

nd

$$a_0\lambda_{\hat{01}} - a_1\lambda_{11} = M_{01} + E[N(t)]E[X(t - T)]$$

vhere

$$\lambda_{00} = E[Y(t)^2] = \lambda_{11},$$

$$\lambda_{01} = E[Y(t)Y(t - T)],$$

$$M_{00} = E[N(t)^2],$$

nd

$$N_{01} = E[N(t)N(t - T)],$$

ve have

$$a_0(\lambda_{00} - \lambda_{01}) = M_{00} + E[X(t)]E[N(t)]$$

nd

$$a_0(\lambda_{01} - \lambda_{00}) = M_{01} + E[N(t)]E(X(t - T)],$$

so that, using stationarity again, $M_{00} = -M_{01}$ and, hence,

$$M_{00} + M_{11} + 2M_{01} = 0,$$

or the second order distribution of $N(t)$, $N(t - T)$ is degenerate, which is contrary to hypothesis. Hence, $C_x(t_0, t_1)$ is Gaussian, and T being arbitrary, all second order distributions are Gaussian. However, since the process is Markoffian, this implies that $X(t)$ is a Gaussian process.

The conditions that the two-point optimal filter be linear for every $T > 0$ is perhaps too stringent and can probably be weakened. Indeed, in the case of discrete-parameter processes (or time series) we can prove a stronger version, as will be shown in Theorem 9.

Theorem 9: Let N_n, X_n, $-\infty < n < \infty$ be strictly stationary discrete parameter processes, let N_n be Gaussian with nondegenerate second-order densities, and let X_n be a Markoff process. Then a necessary and sufficient condition for

$$E[X_n \mid Y_n, Y_{n-1}] = \hat{E}[X_n \mid Y_n, Y_{n-1}] \tag{38}$$

is that X_n be Gaussian also.

Proof: Sufficiency again being trivial and well-known, we need only prove necessity. Here we can make use of Theorem 8, from the proof of which we readily obtain that the joint distribution of X_n and X_{n-1} is Gaussian. But now the joint density of X_n, X_{n-1}, X_{n-2}, because X_n is Markoffian, can be written

$$P[X_n, X_{n-1}, X_{n-2}] = P(X_n \mid X_{n-1})P(X_{n-1} \mid X_{n-2})P(X_{n-2}). \tag{39}$$

Now, since $P(X_n, X_{n-1})$ is Gaussian, so also is the conditional density $P(X_n \mid X_{n-1})$, and, by stationarity, so is $P(X_{n-1} \mid X_{n-2})$. Hence, all the factors in the right side of (39) are Gaussian, and, hence, so is the left side. In a similar manner, it readily follows (by induction, if necessary) that all joint distributions are Gaussian.

Extension of Theorems 8 and 9 to stationary vector Markoff processes is apparent. Indeed, the proof of Theorem 8 shows that for stationary processes, there is at most one characteristic function solution to (30) under the asumption of a nondegenerate second moment matrix for the fixed distribution, so that if there is one other it is automatically unique. Since the second process is usually Gaussian, perhaps Theorem 8 is adequate for most practical purposes.

Conclusions

A new approach to the least-squares optimization theory has been developed. Instead of determining an optimal system for given processes, we characterize the processes for which a prescribed system is optimal. If the designer has a particular type of system in mind, then our results can be used to describe the class of processes for which such a system will be optimal. An immediate advantage of this approach is that the general nonlinear problem now becomes a linear one—although time variant.

In an important subclass of problems, signal and noise are additive and independent and some function of the signal is the desired output. Here we have characterized the class of signal processes for which a given "extraction" system is optimal, assuming that the noise process is known. In some cases it turns out that there is exactly one signal process for a given noise process and this opens the possibility of recognizing the signal by the kind of extraction system that is optimal for it. For stationary nth order Markoff processes, for example, the signal is Gaussian if, and only if, the optimum n-point filter is linear and the noise is Gaussian.

Appendix I

We wish to prove the necessity part of Theorem 4. Thus, we are given that one solution of

$$C_N(t)Q(D)C_x(t) = P(D)[C_N(t)C_x(t)] \tag{40}$$

where

$$Q(D) = \sum_1^n b_k D^k,$$

$$P(D) = \sum_1^n a_k D^k,$$

and

$$C_N(t) = \exp \frac{\lambda_2^2 t}{2}$$

is given by

$$C_x(t) = \exp - \frac{\lambda_1^2 t}{2}.$$

We now have to show that there is no other solution that is a characteristic function with mean zero and variance λ_1^2. We are, of course, assuming $\lambda_1, \lambda_2 \neq 0$, and that we are only after characteristic functions which have a MacLaurin series expansion around the origin. First, let

$$G_1(t) = \exp - \frac{\lambda_1^2 t^2}{2}$$

and

$$G(t) = \exp - \frac{\lambda^2 t^2}{2}, \qquad \lambda^2 = \lambda_1^2 + \lambda_2^2,$$

and, following (42), suppose we express the possible characteristic function as

$$C_x(t) = G_1(t)\left[1 + \sum_3^m \alpha_k (it)^k\right] = G_1(t)\,V(t). \tag{41}$$

That is, the distribution of $X(t_0)$ is expressed by a finite Gram-Charlier expansion.

Let

$$u(t) = \sum_3^m k\alpha_k (it)^{k-1} = \frac{d}{d(it)}\,V(t).$$

Then, substituting (41) into (40), we have

$$C_N(t)\left[\sum_1^n b_i \sum_{k=0}^i C_k^i\, D^{i-k}G_1(t)\, D^k V(t)\right] - \sum_1^n a_i \sum_{k=0}^i C_k^i\, D^{i-k}G(t)\, D^k V(t) = 0$$

where the C_k^i are the binomial coefficients. If we collect the terms in this expression for $k = 0$, we note that by virtue of the fact that $C_x(t) = G_1(t)$ is a solution of (1), these terms already equate to zero. Hence, we have

$$C_N(t)\left[\sum_1^n b_i \sum_1^i C_k^i\, D^{i-k}G_1(t)\, D^{k-1}U(t)\right] - \sum_1^n a_i \sum_1^i C_k^i\, D^{i-k}G(t)\, D^{k-1}U(t) = 0.$$

If we now use

$$D^k[G_1(t)] = (i\lambda_1)^k H_k(t\lambda_1) G_1(t)$$

and

$$D^k[G(t)] = (i\lambda)^k H_k(t\lambda) G(t),$$

this can be expressed as

$$\sum_1^n b_i \sum_1^i C_k^i (i\lambda_1)^{i-k} H_k(t\lambda_1)\, D^{k-1}U(t) - \sum_1^n a_i \sum_1^i C_k^i (i\lambda)^k H_k(t\lambda)\, D^{k-1}U(t) = 0$$

Collecting derivatives of $U(t)$, we have

$$\sum_{k=1}^n \left[\sum_{i=k}^n C_k^i (b_i (i\lambda_1)^{i-k} H_{i-k}(t\lambda_1) - a_i (i\lambda)^{i-k} H_{i-k}(t\lambda))\right] D^{k-1}U(t) = 0. \tag{42}$$

If we substitute

$$U(t) = \sum_{k=3}^m k\alpha_k (it)^{k-1}$$

in this, we obtain an identity in powers of t, and th coefficient of the highest degree, namely $n + m - 2$, i given by omitting nonzero multiplicative constants,

$$b_n \lambda_1^{2n-2} - a_n \lambda^{2n-2}$$

and this must be zero. However, this cannot be zero, sinc for $G_1(t)$ to be a solution we have already seen that w must have

$$b_n \lambda_1^{2n} = a_n \lambda^{2n}.$$

Hence, this proves that $V(t)$ cannot be a polynomial, o $C_n(t)$ cannot have a finite Gram-Charlier expansion. In the more general case, where

$$V(t) = \sum_3^\infty \alpha_k (it)^k$$

we note that, for t small we can express $V(t)$ as

$$V(t) = \alpha_m (it)^m [1 + 0(t)]$$

where α_m is the first nonzero coefficient. Substituting th into (42) again, we can prove again that $V(t)$ must b identically a constant. This proves the necessity part a required.

Bibliography

[1] L. A. Zadeh, "Optimum nonlinear filters," *J. Appl. Phys* vol. 24, pp. 396–404; April, 1953.

[2] A. G. Bose, "A Theory of Nonlinear Systems," Res. Lab. Electronics, Mass. Inst. Tech., Cambridge, Mass., Tech. Rep No. 309; 1956.

[3] A. V. Balakrishnan and R. Drenick, "On optimum nonlinea extraction and coding filters," IRE Trans. on Informatio Theory, vol. IT-2, pp. 166–173; September, 1956.

[4] R. G. Laha, "On a characterization of stable law with finit expectation," *Ann. Math. Statistics*, vol. 27, pp. 187–195; Marc 1956.

[5] L. Doobb, "Stochastic Processes," John Wiley and Sons, Inc New York, N. Y.; 1953.

[6] M. Loeve, "Probability Theory," D. Van Nostrand Co., Inc New York, N. Y.; 1955.

Correction

"On New Classes of Matched Filters and Generalizations of the Matched Filter Concept," David Middleton, these TRANSACTIONS, June, 1960.

The *Editor* wishes to call attention to, and to correct, number of typographical errors in the paper of the above title, which were inadvertently not eliminated in the final stages of printing. These are:

In (3b), lower case $s(t)$ should follow a_0.

Add subscript capital T to $G^{(2)}$ in (5).

On the fifth line of footnote 24, *vis* should be *viz*.

Two lines below (11), c_i should be v_i, and on the line before (12), γ should read $\boldsymbol{\gamma}$.

On the eighth line of the second column on page 353, $V(t)$ should be $v(t)$.

Two lines below (32), delete the comma after "general," and four lines below (32), the second part of the equation should read:

$$\neq [H'(t_i, t_i)\Delta t] \ (\neq 0).$$

Three lines above (33), Q_0' should be $\mathbf{Q}_0'$.

In footnotes 38 and 41, the page reference on the last line should be p. 412 instead of p. 348.

In (37b), r_i should be v_i.

Three lines above (43), the expression $h(t_i, t_i)$ should be $H(t_i, t_i)$.

The right side of (48) should read $H'(t, \tau) \neq H'(\tau, t)$.

In (53), the term $\langle \psi_T \rangle$ should be $\langle \Psi_T \rangle$.

On the tenth line of footnote 44, P_C should be ρ_C.

Three lines below (57), λ_n should be λ_m.

Eleven lines below (57), replace the first 2 in $h_M(T - x) = 2 \cos [2\pi m(T - x)/T]$ by ϵ_m, the Neumann factor $\epsilon_0 = 1$, $\epsilon_m = 2$, $m \neq 0$.

On the ninth line below (58b), $e^{i\phi}T$ should be $e^{i\phi\tau}$.

In the column entitled "Filter Structure" in Table I, the equation for b) of 3) should be $H'(t, \tau) \neq H'(\tau, t)$.

Correspondence

Remarks on Sine Waves Plus Noise*

In a recent letter,[1] Levine and McGhee presented a short table of the first order cumulative distribution functions (*cdf*) of a sine wave of random phase plus Gaussian noise. Their table is computed from an integral given by S. O. Rice.[2] However, in a later publication,[3] Rice gave explicit expressions for the first order probability density functions (*pdf*) and *cdf* in terms of series of Hermite functions and error functions. From Cramer's inequality, Rice's series are seen to converge more rapidly than the series for exp $(r^2/2^{1/2})$, where r is the signal/noise ratio. Thus, a few terms would suffice for moderate values of r. For example, 20 terms would give a 5-place accuracy for $r < 2.5$. In addition, Rice gave several asymptotic formulas for the first order *pdf* for large r.

The writer recently derived exact expression both for the first and for the second order *pdf*'s of the previously mentioned stationary process in terms of rapidly converging series in (tabulated) Bessel functions of purely imaginary argument. The results are as follows:[4]

For the first order *pdf*,

$$1)\qquad p_1(y) = (2\pi)^{-1/2}\exp\left(-\frac{y^2}{2}\right)\cdot\exp\left(-\frac{r}{2}\right)\sum_{q=-\infty}^{\infty}(-1)^q I_q\left(\frac{r}{2}\right)\cdot I_{2q}((2r)^{1/2}y)$$

where

$$y = \frac{\text{signal} + \text{noise}}{(\text{noise power})^{1/2}},$$

$$r = \frac{\text{signal power}}{\text{noise power}}.$$

For the second order *pdf*,

$$2)\qquad p_2(y_1, y_2) = (2\pi)^{-1}(1-\rho_\Delta^2)^{-1/2}\cdot\exp\left[-\frac{y_1^2+y_2^2-2\rho_\Delta y_1 y_2}{2(1-\rho_\Delta^2)}\right]\cdot\exp\left[\frac{-r(1-\rho_\Delta\cos\omega\Delta)}{1-\rho_\Delta^2}\right]\cdot\sum_{q=-\infty}^{\infty}\cos\left\{2q\tan^{-1}\left[\frac{(y_1-y_2)(1+\rho_\Delta)}{(y_1+y_2)(1-\rho_\Delta)}\tan\frac{\omega\Delta}{2}\right]\right\}\cdot I_q\left[\frac{r(1-\rho_\Delta\cos\omega\Delta)}{1-\rho_\Delta^2}\right]\cdot I_{2q}\left[\left\{2r\left\{\left[\frac{(y_1-y_2)\sin\frac{\omega\Delta}{2}}{1-\rho_\Delta}\right]^2+\left[\frac{(y_1+y_2)\cos\frac{\omega\Delta}{2}}{1+\rho_\Delta}\right]^2\right\}\right\}^{1/2}\right]$$

where Δ is the time interval, ρ_Δ the noise autocorrelation and ω the signal frequency. The use of Rice's series or that given in 1) would appear to be more convenient than numerical integration for evaluation of the first order distribution functions in some cases.

R. Leipnik
Michelson Lab., USNOTS
China Lake, Calif.

*Received by the PGIT, July 8, 1960.

[1] A. Levine and R. B. McGhee, "Cumulative distribution functions for a sinusoid plus Gaussian noise," IRE Trans. on Information Theory, vol. IT-5, pp. 90–91; June, 1959.

[2] S. O. Rice, "Mathematical Analysis of Random Noise," Bell Telephone Sys. Monograph No. B-1559, p. 105; 1945.

[3] S. O. Rice, "Statistical properties of a sine wave plus random noise," *Bell Sys. Tech. J.*, vol. 27, pp. 109–157; January, 1948.

[4] The results are derived at length in an article by the author to appear in *Z. Angew. Math. u Phys., Zurich.*

Correction to a Paper by D. G. Lampard*

As a consequence of a recent communication from D. G. Lampard of Sidney, Australia, the author wishes to make a correction in his recent paper[1] in which the following sentence appears:

> Eq. (4) does not seem to have been observed before except by Levin[5] whose formula is in error.

Mr. Lampard kindly points out that (4) has been observed before by several other authors, including himself. He lists a set of references.[2]

I. S. Reed
Lincoln Lab.
Mass. Inst. Tech.
Lexington, Mass.

* Received by the PGIT, February 18, 1960.

[1] I. S. Reed, "On the use of Laguerre polynomials in treating the envelope and phase components of narrow-band Gaussian noise," IRE Trans. on Information Theory, vol. IT-5, pp. 102–105; September, 1959.

[2] D. G. Lampard and J. F. Barrett, "An expansion for some second-order probability distributions and its application to noise problems," IRE Trans. on Information Theory, vol. IT-1, pp. 10–16; March, 1955.

K. S. Miller, R. I. Bernstein, and L. E. Blumenson, "Generalized Rayleigh processes," *Quart. Appl. Math.*, vol. 16; July, 1958.

M. Nakagami, K. Tanaka, and M. Kanchisa, "The *m*-distribution as the general formula of intensity distribution of rapid fading," *Mem. Fac. Engrg. Kobe Univ., Japan*; March, 1957.

S. O. Rice, "Communication in the presence of noise-probability of error for two encoding schemes," *Bell Sys. Tech. J.*, vol. 29; January, 1950.

Note on "On Upper Bounds f Error Detecting and Error Correcti Codes of Finite Length"*

In a recent article,[1] Wax cites the resu of Laemmel that the best value actual found for the number of sequences length 12 with a minimum Hamming d tance of five is 24. Here it will be illustrat how this number has been increased 28. The latter number is in closer agre ment with Wax's upper bound of 46 than the former.

By using all of the distinct cyclic perm tations, the following pair of binary s quences of length 12 will produce sequences with a minimum distance five between any two sequences:

1 1 0 0 0 0 1 0 1 1 0 0
1 1 1 1 0 1 0 0 1 0 0 0.

Adding to these the sequence of all zero the sequence of all ones, and the two di tinct cycles of 0 1 0 1 0 1 0 1 0 1 0 1 gi a total of 28 sequences. These sequenc were also run off on a computer in a attempt to increase their number. Ho ever, no new sequences were uncovere

Additional pairs of sequences of leng 12 which autocorrelate and crosscorrela favorably are:

0 0 1 1 1 0 0 1 0 0 1 0
0 0 0 0 1 0 1 1 0 1 1 1

0 0 1 1 1 0 0 1 0 0 1 0
1 1 0 1 1 0 1 0 1 0 0 0

1 1 0 0 0 0 1 0 1 1 0 0
0 1 0 0 0 1 0 0 1 1 1 1.

R. G. Fry
Amherst Engineering La
Sylvania Electric Products, In
Buffalo, N.

* Received by the PGIT, March 3, 1960.

[1] N. Wax, "On upper bounds for error detecti and error correcting codes of finite length," I Trans. on Information Theory, vol. IT-5, 168–174; December, 1959.

A Note on Single Error Correcti Binary Codes*

In a recent paper on double adjace error correcting codes,[1] the possibility deriving a class of related binary sing error correcting codes was mention

* Received by the PGIT, February 15, 19

[1] N. Abramson, "A class of systematic codes non-independent errors," IRE Trans. on Inf mation Theory, vol. IT-5, pp. 150–157; Decemb 1959.

arenthetically. It now appears that this ass of codes may have some advantages simplicity of equipment over the ordinary amming[2] codes. In this note, therefore, e shall explain in more detail the conruction and properties of these codes.

Consider single error correcting block odes with k information digits and $- k = r$ check digits. It is well-known[2] iat for a given r, the maximum value of n just $2^r - 1$. We shall restrict ourselves codes where n is equal to this maximum alue. Let $a_1a_2 \cdots a_n$ be the binary digits omprising a word of such a code. Then, r a Hamming code, the a_i must satisfy ie (mod 2) equation.

$$a_1 + x_2a_2 + \cdots + x_na_n = 0 \quad (1)$$

here x_1 is the $r \times 1$ matrix

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

is the matrix

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

is the matrix

$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

nd the remaining x_i are developed in rder from the binary counting sequence ntil x_n, which is given by

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

For the class of codes mentioned by bramson,[1] however, the a_i must satisfy ie (mod 2) equation[3]

$$y_1a_1 + y_2a_2 + \cdots + y_na_n = 0, \quad (2)$$

where

y_1 is the $r \times 1$ matrix

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (3)$$

$$y_2 = Ty_1,$$

$$y_3 = T^2y_1,$$

$$\vdots$$

$$y_n = T^{n-1}y_1,$$

and T is any r by r binary matrix whose binary characteristic polynomial is both irreducible and of a maximal period.[4] An alternate characterization is that the elements of the top row of $y_1, y_2, \cdots y_n$ define an m-sequence[5] ending in $r - 1$ zeros and the elements of the j^{th} row of the y_i define the same m-sequence shifted to the right by $j - 1$ digits.

We shall illustrate the preceding paragraph for the case $n = 7$ and $k = 4$. An acceptable T matrix for any $r \leq 19$ (and thus, $n \leq 524{,}287$) may be obtained from Marsh's tables.[6] We select

$$T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (4)$$

so that (2) may be written as

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} a_1 + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} a_2 + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} a_3 + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} a_4$$

$$+ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} a_5 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} a_6 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} a_7 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (5)$$

Note that for this code, it is possible to take the last three digits a_5a_6 and a_7 as the parity digits, and the first four digits a_1, a_2, a_3 and a_4 as the information digits. It can be shown in general that the choice of y_1, as in (3), allows one to place all parity digits at the end of the block—an important advantage in many applications. Furthermore, this may be done without destroying the simplicity of equipment necessary to implement the code. Since the T matrix corresponds to a time delay of one unit in an r-stage feedback shift register, this shift register may be used to time the generation of the parity digits.[7] For example, Fig. 1 shows the shift

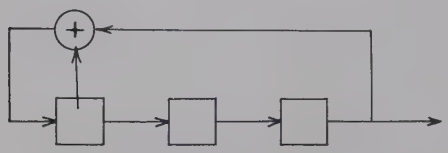


Fig. 1.

register corresponding to (4) which generates the m sequence used in (5), 1110100.

Implementations based on the use of the feedback shift register have also been obtained by Peterson[8] and Meggitt.[9]

N. M. Abramson
Stanford University
Stanford, Calif.

[2] R. W. Hamming, "Error correcting and error tecting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 47–160; April, 1950.

[3] The notation used in (2) was originated by E. Meggitt in "Error correcting codes for correct- g bursts of errors," to be published in the *IBM ournal.*

[4] B. Elspas, "Theory of autonomous linear sequential networks," IRE Trans. on Circuit Theory, vol. CT-6, pp. 45–60; March, 1960.

[5] N. Zierler, "Several Binary-Sequence Generators," Lincoln Lab., Mass. Inst. Tech., Lexington, Mass., Tech. Rept. No. 95.

[6] R. W. Marsh, "Table of Irreducible Polynomials Over $GF(2)$ Through Degree 19," National Security Agency, Washington, D. C.; October 24, 1957.

[7] N. Abramson and B. Elspas, "Double-error-correcting encoders and decoders for non-independent binary errors," *Proc. UNESCO Conf. on Information Processing*, International Documents Service, Columbia University Press, New York, N. Y.; 1960.

[8] W. Peterson, "Error Correcting and Error Detecting Codes," to be published by Technology Press.

[9] J. Meggitt, "Error correcting codes for correcting bursts of errors," to be published in the *IBM J. Res. & Dev.*

Transmission of Photographic Data by Electrical Transmission*

The purpose of this note is to dispel some misconceptions arising from the following statement in the Space Handbook:[1]

> "A communication system with a bandwidth of 6 megacycles per second will have to operate continuously for 22.5 minutes to transmit the quantity of information that can be stored on a single 9 × 9 inch photograph at 100 lines per millimeter."

The 22.5-minute transmission time can be deduced from any of a large set of plausible assumptions. However, the transmission time is quite sensitive to the assumptions which are made, and the statement in the Space Handbook should not be accepted without reservation. The authors have found two citations of this statement (in classified documents) which show in each case that the author has been misled.

As an example, here is a set of assumptions which will lead to the 22.5-minute figure: suppose that the source material

* Received by the PGIT, February 10, 1960.

[1] "Space Handbook: Astronautics and Its Applications, Staff Report of the Select Committee on Astronautics and Space Exploration," U. S. Government Printing Office, Washington, D. C., 86th Congress, 1st Session, House Document No. 86, p. 181; 1959.

consists of $(200)^2$ independent points per square millimeter, each carrying four bits of information (*i.e.*, 16 distinguishable gray levels), and that the electrical transmission system transmits information at the rate of one bit per second for each cycle of bandwidth.

These assumptions may be valid in a particular case. In other cases, the information content of the picture may be less and the channel capacity of the transmission link greater. Consider the following example.

A photographic resolution of N lines per millimeter means[2] that an array of $2N$ equally-spaced lines, alternately black and white, can be resolved. This does not mean, in general, that 16 gray levels can be distinguished at this resolution; on the contrary, the resolution is defined in terms of the number of lines per millimeter at which *only two gray levels* can be distinguished.

Generally speaking, a picture does not preserve perfect clarity of tonal detail right up to the limit of resolution. On the contrary, the detail "cuts off" with decreasing wavelength. If the resolution is determined, for example, by the finite size of a scanning aperture, the loss of gray-scale resolution as a function of wavelength can be computed explicitly.

To show how this affects the information content of the picture, imagine the following idealized case: for long wavelengths, the number of gray levels is G, and for short wavelengths, the number of gray levels is proportional to the Nth power of wavelength, *i.e.*, λ^n or $6n$ db per octave cutoff. This results in the signal spectrum of Fig. 1. The figure is plotted for a maximum number of gray scales G_m of 16, a resolution $1/\lambda$ of 100, and a cutoff rate n of 1, or 6 db per octave. The information per spot is proportional to

$$\overline{\log_2 G}$$

$$= \lambda \int_0^{1/\lambda} \log_2 (G + 1)\, d(1/\lambda)$$

$$\simeq \lambda \int_0^{1/\lambda} \log_2 G\, d(1/\lambda)$$

$$= n \log_2 (1 - G^{-1/n})$$

$$= 1.35 \quad \text{for} \quad G_m = 16, \qquad n = 1,$$

$$= 2.16 \quad \text{for} \quad G_m = 16, \qquad n = 2,$$

$$= 2.60 \quad \text{for} \quad G_m = 16, \qquad n = 3.$$

The figures are quite insensitive to G_m. In order to show how much is taken away by the cutoff, the curve is plotted with a linear frequency scale as well as with the customary logarithmic scale. An average value of 2.2 bits per point seems on the whole more realistic than 4.

[2] D. G. Fink, "Television Engineering," McGraw-Hill Book Co., Inc., New York, N. Y., 2nd ed., p. 24 ff.; 1952.

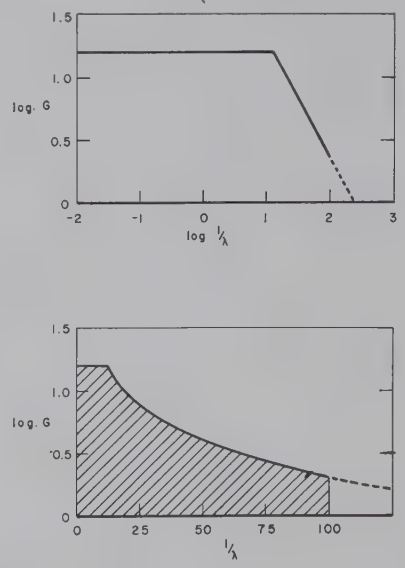


Fig. 1—Number of gray levels as a function of wavelength.

Now let us look at the transmission link. Many types of 6-mc links are conceivable, but let us choose as an example a link which might be designed to transmit commercial television with a 26-db SN ratio at low frequencies and a 4-mc bandwidth. Assume that the channel has a 3-db point at 4 mc and that it cuts off thereafter at 36 db per octave. Transmission is reduced 11 db at 6 mc, leaving an available SN ratio greatly in excess of the 3 db required at the cutoff of the picture information. (A bit of shaping, either pre- or postemphasis, is required to preserve the *appearance* of the picture, but not to preserve the information.)

With this channel, a line could be scanned at a rate

$$v = \frac{6 \times 10^6}{N} \quad \text{mm/second}$$

without significant loss. The number of scanning lines required is not $2N$/mm, as one might imagine, but[3] $2\sqrt{2}N$, or in practice, $3N$. Hence the time required to scan a picture of area A square inches is

$$t = (25.4)^2 A \cdot (N/B) \cdot 3N$$

$$= \frac{1940 A N^2}{B} \quad \text{or, say,} \quad 2000 A N^2/B,$$

where $B = 6 \cdot 10^6$ is the bandwidth. For

N = 100 lines/mm,

A = 81 square inches,

t = 262 seconds = 4.4 minutes.

Countless examples could be generated, each leading to a different result.

In conclusion, one can say that the estimate given in the Space Handbook is extremely conservative and should not be construed as a limit on any particular communication system without looking further to see whether the assumptions which it is based are valid in the case hand. A 6-mc channel which will transm a 9 × 9-inch photograph with a resoluti of 100 lines/mm in four minutes is who practical.

G. Raisbe
Inst. for Defense Analys
AR
Washington, D.
J. Goldhamm
Chicago Aerial Industri
Barrington, I

[3] *Ibid.*, pp. 25–26.

A Note of Caution on the Squar Law Approximation to an Optimu Detector*

When zero-mean Gaussian noise volta is added to a sinusoidal signal, the envelo of the sum has the well-known[1] probabili density function

$$W(R;\, A) = \begin{cases} \dfrac{R}{\psi} e^{-(A^2+R^2)/2\psi} I_0\left(\dfrac{AR}{\psi}\right), & R \geq 0 \\ 0, & R \leq 0 \end{cases}$$

in which A is the amplitude of the sign sinusoid, ψ is the mean square of the nois the random variable R is the amplitu of the envelope and $I_0(v)$ is the modifi Bessel function of order zero and arg ment v.

In terms of normalized quantities, becomes

$$w(x;\, a) = \begin{cases} xe^{-(a^2+x^2)/2} I_0(ax), & x \geq 0 \\ 0, & x \leq 0 \end{cases}$$

in which

$$x = R/\sqrt{\psi}, \qquad a = A/\sqrt{\psi}.$$

For small signal-to-noise ratios, it h been common in the past to approxima the optimum detector associated with t probability density function by a squa law detector.[2,3] The optimum detec means here a device whose functional fo is such that its output represents t logarithm of the ratio of the appropri *a posteriori* probabilities.

* Received by the PGIT, March 3, 1960; revis manuscript received, April 12, 1960.

[1] S. O. Rice, "Mathematical analysis of rand noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–3 June, 1944; vol. 24, pp. 46–156, January, 19

[2] D. Middleton, "Statistical criteria for the tection of pulsed carriers in noise (pts. I and II *J. Appl. Phys.*, vol. 24, pp. 371–378, 379–391; Ap 1953.

[3] W. W. Peterson, T. G. Birdsall, and W. Fox, "The theory of signal detectability," I Trans. on Information Theory, vol. IT-4, 171–212; September, 1954.

Bussgang and Middleton[4,5] and Blasbalg[6] ve pointed out that this square-law proximation can lead to certain pitfalls nen improperly applied. In the Russian erature, Flejshman,[7] in a paper devoted the subject, gives a detailed discussion the same problem under what are really e same assumptions. Still, we find that is problem is not fully appreciated and at the simple square-law approximation often accepted without qualifications. appears important once again to draw tention to a difficulty which often arises. Let z be the logarithm of the ratio of *posteriori* probabilities

$$z = \ln \frac{w(x; a_1)}{w(x; 0)} \tag{3}$$

which x is the observed value of the velope and the amplitude a_1 is the one osen to represent the hypothesis that e signal is present. Substituting from) in (3), one gets

$$z = -\frac{a_1^2}{2} + \ln I_0(a_1 x). \tag{4}$$

he optimum detector in problems in- lving signal plus noise has a law which functionally of the form (4), where x d z are, respectively, the normalized velope at the input and the normalized ltage at the output of the detector. In the region of the detector law where e instantaneous envelope is small ($x \ll 1$), (4) can be approximated. The problem hinges on keeping enough terms in the Taylor series expansion of the logarithm. For small $a_1 x$, one gets

$$z = -\frac{a_1^2}{2} + \frac{a_1^2 x^2}{4} - \frac{a_1^4 x^4}{64} + 0(a_1^6 x^6). \tag{5}$$

It is hasty to assume from (5) that the optimum detector law contains just the first two terms even when the envelope x and a_1 are both small. The term in x^4 must be appropriately included. To demonstrate this, consider $E(z)$, the expected value of z. Since from (2) it can be shown that

$$\begin{aligned} E(x^2) &= 2 + a^2 \\ E(x^4) &= 8 + 8a^2 + a^4 \end{aligned} \tag{6}$$

keeping the term in x^4 in (5), it follows that

$$E(z) = -\frac{a_1^4}{8} + \frac{a_1^2}{4}\left(1 - \frac{a_1^2}{2}\right)a^2 - \frac{a_1^4}{64}a^4 + \cdots . \tag{7}$$

For small a_1 *and* a, (7) is approximately

$$E(z) \doteq \frac{a_1^2}{4}\left(a^2 - \frac{a_1^2}{2}\right). \tag{8}$$

However, if only the first two terms of (5), up to x^2, had been kept then one would have

$$z' = -\frac{a_1^2}{2} + \frac{a_1^2 x^2}{4}, \tag{9}$$

and one would obtain

$$E(z') = \frac{a_1^2}{4}a^2. \tag{10}$$

Thus, the "square-law" approximation of $E(z)$ for $a = 0$ would have been taken improperly as 0, rather than correctly as $-a_1^4/8$. This error in $E(z)$ which results from (9) is not merely quantitative; it implies that the expected output of the optimum detector corresponding to (4) is never negative, no matter how small a becomes.

In the case of sequential detection[4,5] the magnitude of $E(z)$ is inversely proportional to the average sample size. Thus a value of zero for $E(z)$ could falsely imply infinite sample sizes.

It is possible that other results obtained by the use of the square-law approximation (9) should be re-examined, for they too may be affected.

One way out of the difficulty consists of replacing $-a_1^4 x^4/64$ in (5) by its expected value $-a_1^4/8$ in the weak signal case. This replacement is equivalent to a change of bias, leaving the detector still a square-law device. Such an approach removes the basic difficulty[4,5,8] of reconciling (8) and (10).

Even keeping the term in x^4 in (5) may not suffice when a_1 is large compared to unity (a_1 is the hypothesized voltage signal-to-noise ratio), even though a (the actual signal-to-noise ratio) itself is small. Enough terms in a_1 must be kept in (7) to evaluate the coefficients of a^0 and a^2.

Of course, if a detector were chosen to follow the square law represented by the first two terms in (5), the results questioned here would apply. However, such a detector is not optimum in the sense of (3).

The square-law approximation (9) is so often used without attention to the critical influence of higher-order terms that we have felt it important to offer this note of caution.

J. J. Bussgang
W. L. Mudgett
RCA
Burlington, Mass.

[4] J. J. Bussgang and D. Middleton, "Sequential etection of Signals in Noise," Cruft Lab., Harvard niversity, Cambridge, Mass., Tech. Rept. No. 175; ugust 31, 1955.

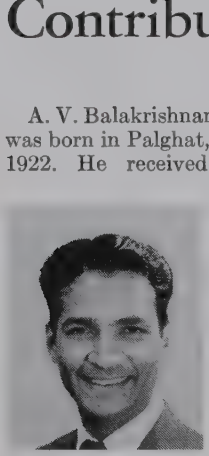
[5] J. J. Bussgang and D. Middleton, "Optimum quential detection of signals in noise," IRE Trans. Information Theory, vol. IT-1, pp. 5–18; ecember, 1955.

[6] H. Blasbalg, "The sequential detection of a ne-wave carrier of arbitrary duty factor in Gaussian ise," IRE Trans. on Information Theory, vol. T-3, pp. 248–256; December, 1957.

[7] B. S. Flejshman, "On the optimal detector th a log I_0 characteristic for the detection of a eak signal in the presence of noise," *Radiotekhnika Elektronika*, vol. 2, pp. 726–734; June, 1957. (In ussian.)

[8] D. Middleton, "An Introduction to Statistical Communication Theory," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 836, 876, 900; 1960.

Contributors


A. BALAKRISHNAN

A. V. Balakrishnan (S '43—A '55—M '56) was born in Palghat, India, on December 4, 1922. He received the Bachelor's and Master's degrees in physics from the University of Madras, India, in 1945. He came to the United States in 1947 on a two-year Indian Government scholarship. In 1950 he was awarded the Master's degree in electrical engineering and in 1954, the Ph.D. degree in mathematics, both from the University of Southern California, Los Angeles.

While doing graduate work, he was a laboratory assistant, teaching associate and lecturer at the University of Southern California, and an assistant instructor in the Mathematics Department of Yale University, New Haven, Conn.

From 1954 to 1956, he was with RCA, Camden, N. J., working on communication and control problems, including multipath transmission of video signals, noise cancellation systems and nonlinear filters. From 1956 to 1957 he was an assistant professor of mathematics at the University of Southern California, and from 1957 to 1959 at the University of California, Los Angeles. At present he is with The Space Technology Laboratories, Los Angeles, heading a research group on communication and control theory.

Dr. Balakrishnan is a member of Tau Beta Pi and Sigma Xi.

❖


M. BLUM

Marvin Blum (M '56) was born on June 18, 1928, in New York, N. Y. He received the B.S. degree from Brooklyn College, N. Y., in 1948, and has taken graduate courses in mathematics, physics, and electrical engineering at George Washington University, Washington, D. C., American University, Washington, D. C., Maryland University, College Park, Md., the National Bureau of Standards School, Denver, Colo., and the University of California, Los Angeles Extension.

He worked at the National Bureau of Standards in the Central Radio Propagation Laboratory in Washington, D. C., until 1950. He then transferred to the Ordinance Division, where he conducted radar reflection studies relating to missile proximity fuzes. In 1954, he was employed by Convair, San Diego, Calif., where he conducted theoretical investigations in smoothing and prediction filters, noise simulation, and data reduction. He left the Convair Astronautics Division in July, 1959, to join the System Development Corp., Santa Monica, Calif., where he is working on prediction and smoothing filters as applied to space defence.

Mr. Blum is a member of the Society for Applied Mathematics.

❖


E. N. GILBERT

E. N. Gilbert was born in Woodhaven, N. Y., on July 25, 1923. He received the B.S. degree in physics from Queens College, Flushing, N. Y., in 1943, and the Ph.D. degree in mathematics from the Massachusetts Institute of Technology, Cambridge, in 1948. At M.I.T. he held an Applied Mathematics Fellowship.

From 1944 to 1946 he designed antennas at the M.I.T. Radiation Laboratory. Since 1948 he has been a member of the Mathematical Research Department of Bell Telephone Laboratories in Murray Hill, N. J., where his current main interests are combinatorial analysis and probability and their applications to problems in switching and coding.

Dr. Gilbert is a member of the American Mathematical Society.

❖

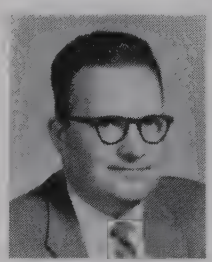

W. A. JANOS

William A. Janos (M '59) was born on November 9, 1926, in Easton, Pa. He received the B.S. degree in physics from Rutgers University, New Brunswick, N. J., in 1951, and the M.A. and Ph.D. degrees in 1954, and 1958, respectively, both in physics, from the University of California, Berkeley. He was recipient of the University's appointed teaching assistantship in the physics department, and also the Convair Scholarship Award.

He served in the U. S. Army from 1945 to 1947. From 1951 to 1960, he was with Convair and Convair-Astronautics, San Diego, Calif., where he engaged in applied analysis and spectral theory related to analytical dynamics, wave propagation and diffraction, variational techniques in least-time trajectories for thrust-propelled flight, control system analysis and synthesis, noise theory and optimal linear estimatio He is presently a staff physicist in t Physics Department of the Advanced D velopment Laboratory, Raytheon C Wayland, Mass.

Dr. Janos is a member of the Americ Physical Society.

❖


M. V. JOHNS, JR.

M. Vernon Johns, Jr., was born Berkeley, Calif., on September 27, 192 He received the B.A. degree in economi from Stanford U versity, Stanfor Calif., in 1949 a the Ph.D. degree mathematical stat tics from Columb University, Ne York, N. Y., in 195 Since 1956 he h been with Stanfo University, first as research associate statistics and, sin 1957, as assista professor of statistics. His research activi has been mainly in the areas of statisti decision theory and probability proble related to statistics.

Dr. Johns is a member of the Americ Mathematical Society, the Institute Mathematical Statistics, the Americ Statistical Association, the American A sociation for the Advancement of Scie and Sigma Xi.

❖

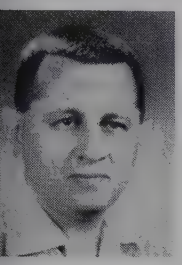

W. W. PETERSON

W. Wesley Peterson (S '49—A '52— '58) was born in Muskegon, Mich., on Ap 22, 1924. He attended the University Michigan, Ann bor, receiving A.B. degree in ma ematics in 1948, M.S.E. degree 1950, and Ph.D. gree in electrical gineering in 1954.

From 1951 to 1 he was a resea associate in the gineering Resea Institute of the U versity of Michig He was employed by the IBM Engineer Laboratory in Poughkeepsie, N. Y., fr 1954 to 1956. In 1956 he joined the facu of the University of Florida, Gainesvi where he is now an associate profes He is currently on leave as visiting associ professor of electrical engineering at Massachusetts Institute of Technolo Cambridge.

Dr. Peterson is a member of the Ameri Mathematical Society and Sigma Xi.

Morris Plotkin was born in Philadelphia, ., on February 9, 1914. He received the S. degree in electrical engineering in 1934, the M.A. degree in mathematics in 1951, and the M.S. degree in electrical engineering in 1952, all from the University of Pennsylvania, Philadelphia.



M. Plotkin

Prior to World War II, he worked in electrical power engineering. Upon discharge from the Navy in 1946, he ned the research staff of the Moore hool of Electrical Engineering at the iiversity of Pennsylvania.

In 1951 he joined the Naval Air Develop-nt Center at Johnsville, Pa., where he rved as a mathematics consultant and rected the operation of digital and analog mputers. He was instrumental in the sign and operation of the first modern rge-scale simulation programs on an ana-g computer. Later, he supervised the tegration of a human centrifuge and an alog computer into a facility for closed-p simulation of aircraft control systems. Since May, 1959, he has been chief of alysis at Auerbach Electronics Corpora-n, Philadelphia, Pa., where he is engaged the mathematical analysis of physical d organizational systems. His current tivities include the development of ecial techniques for radar data extraction d solution of queueing problems in gital communications networks.

Mr. Plotkin is a member of Sigma Xi d the American Mathematical Society.

❖

Richard A. Silverman (M '54—SM '58) as born on June 29, 1926, in Boston, ass. He received the A.B. degree from Harvard University, Cambridge, Mass., in 1946, the M.A. degree from Columbia University, New York, N. Y., in 1948, and the Ph.D. degree from Harvard in 1951.

. A. Silverman

For three years he was associated with the Massachusetts Institute of Technology, Cambridge, t as a staff member of the Lincoln boratory and then as a research associate in the Department of Engineering, Electrical. Currently, he is a research scientist at the New York University Institute of Mathematical Sciences, New York, in the Division of Electromagnetic Research.

Dr. Silverman is a member of Phi Beta Kappa, Sigma Xi, and the Society for Industrial and Applied Mathematics.

❖

Thomas E. Stern (S '54—M '57) was born in New Rochelle, N. Y., on March 29, 1930. He attended the Massachusetts Institute of Technology, Cambridge, where he received his undergraduate education under the cooperative program in electrical engineering. After receiving the B.S. and M.S. degrees in 1953, he became a research assistant at the M.I.T. Research Laboratory of Electronics, and in 1956 received the Sc.D. degree from M.I.T.

T. E. Stern

At present, he is an assistant professor of electrical engineering at Columbia University, New York, N. Y. His areas of research include analog computation, nonlinear network theory and information theory.

Dr. Stern is a member of Eta Kappa Nu and Sigma Xi.

❖

John B. Thomas (S '52—M '56) was born in New Kensington, Pa., on July 14, 1925. He received the A.B. degree from Gettysburg College, Gettysburg, Pa., in 1944, the B.S. degree from The Johns Hopkins University, Baltimore, Md., in 1952, and the M.S. and Ph.D. degrees from Stanford University, Stanford, Calif., in 1953 and 1955, respectively.

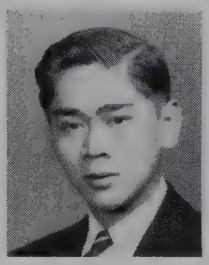
J. B. Thomas

During World War II he served in the U. S. Army. From 1946 to 1952 he was employed by Koppers Co., Inc., Baltimore, Md., in the Industrial Gas Cleaning Department, first as an electrical engineer and then as assistant chief engineer of the department. Since 1955 he has been a member of the Department of Electrical Engineering at Princeton University, Princeton, N. J., and is currently an associate professor in that department.

Dr. Thomas is a member of the American Institute of Electrical Engineers, Tau Beta Pi, and Sigma Xi.

❖

Eugene Wong (S '57) was born in Nanking, China, on December 24, 1934. He received the B.S.E., A.M., and Ph.D. degrees from Princeton University, Princeton, N. J., in 1955, 1958, and 1959, respectively. From 1955 to 1956 he was employed by I.B.M., Poughkeepsie, N. Y., where he engaged in semiconductor research. For the year 1959–1960, he is a National Science Foundation postdoctoral fellow at the University of Cambridge, Cambridge, England.



E. Wong

Dr. Wong is a member of Sigma Xi.

❖

Neal Zierler was born on September 17, 1926, in Baltimore, Md. He received the A.B. degree in physics at the Johns Hopkins University, Baltimore, in 1944. After service in the Navy, he returned to Johns Hopkins for graduate work in physics and then attended Harvard, University, Cambridge, Mass., where he received the M.A. degree in mathematics in 1949 and the Ph.D. degree in 1959.

N. Zierler

He worked at the Ballistic Research Laboratories, Army Ordnance, Aberdeen, Md., in 1951, and at the Instrumentation Laboratory of Massachusetts Institute of Technology from 1952 to 1953. Since April, 1954, he has been a member of the staff of M. I. T.'s Lincoln Laboratory, at Lexington, Mass.

Dr. Zierler is a member of the American Mathematical Society and the Mathematical Association of America.

Book Reviews

Information Theory and Statistics—S. Kullback. (John Wiley and Sons, Inc., New York, N. Y., 1959; and Chapman and Hall, Ltd., London, England; 1959. 395 pages + xvii pages. $12.50.)

This book is about statistics; in particular, the testing of statistical hypotheses. It is not concerned with the sort of applications of information theory to communications problems that constitute the major part of these TRANSACTIONS. Since, however, statistics has to do with gathering and using information (in the wide sense), and since "information theory" (in the Shannon sense) purports to be a quantitative theory of information and its handling, one is tempted to suppose that statistics could be given an information theoretic basis. This book represents a more or less unified account of the author's pioneering efforts in this direction.

Starting with a few fairly simple postulates of an information theoretic nature, but quite *ad hoc* from the statistician's point of view, the author derives an impressive array of old and new results in statistical hypothesis testing and related ideas. Most of the material is derived from two basic notions: 1) the directed divergence from one simple hypothesis to another and 2) the directed divergence from a sample to a hypothesis. Detailed definitions of these concepts will be found in the first five chapters of the book.

By using the tools mentioned above, the author gives, in Chapters 6–13 (comprising 244 pages), detailed and often numerical discussion of examples in multinomial and Poisson populations, contingency tables and multivariate normal populations. Substantial problem lists appear at the end of each chapter. Referencing is extremely thorough in all sections, including problem sections.

In criticism, the reviewer remarks that the reader is unnecessarily exposed to much secondary material before the author's main thesis is reached in Chapter 5; the definition of critical region on page 85 is not clear; in several places, discussion of tests based on the symmetrized divergence could well have been omitted. These are minor faults, however.

The author states in the Preface, "Applications to more general stochastic processes, including sequential analysis will make a natural sequel, but are outside the scope of this book."

S. P. LLOYD
Bell Telephone Labs.
Murray Hill, N. J.

The Theory of Optimum Noise Immunity—V. A. Koteln'ikov (translated from the Russian by R. A. Silverman). (McGraw-Hill Book Co., Inc., New York, N. Y.; 1959. 140 pages + xi pages. Illus. $7\frac{1}{4}$ x $10\frac{1}{4}$. $7.50.)

Vladimir Aleksandrovich Kotel'nikov is one of the few electronics engineers ever to be elected Academician in the Academy of Sciences of the U. S. S. R. In 1947 he published a doctoral dissertation which was subsequently republished in book form in Russian. R. A. Silverman's excellent translation of this Russian classic should be of interest to English speaking communications engineers for both technical and historical reasons.

Kotel'nikov deals with the problem of analyzing the performan of various communication systems in the presence of additiv Gaussian noise. This performance is measured by minimum pro ability of error in the case of discrete messages and minimu mean-square error in the continuous case. The organization of th little book is as neat and concise as Kotel'nikov's mathematic Part I serves to set the framework of the problems considered ar to sharpen the few elementry mathematical tools used in the rest the book. These tools consist of : 1) the expansion of signals defin in the interval $[-T/2, T/2]$ in terms of a set of orthonormal function 2) the use of this expansion to derive the geometrical properti of a signal space, and 3) the corresponding expansion of shot noi and the statistical properties of such noise in signal space. Part then uses these tools to treat the transmission of discrete messag as an hypothesis testing problem. Part III deals with the analys of the transmission of a continuous parameter value treated as statistical estimation problem. In terms of what we now kno much of Part III may be generalized. This does not, howeve detract from the usefulness of Kotel'nikov's analysis. Finall Part IV deals with the extension of the techniques of statistic estimation to cover the analysis of the transmission of continuou time functions. Most of the results in Part IV are restricted to th case when the signal-to-noise ratio is high.

Each of the last three parts begins with a general analysis of th problem and a derivation of the optimum method of detectio After this, the author presents a large number of examples of pra tical communication systems to illustrate the theory. In son cases, the examples are compared with a suboptimum method detection. Each part is concluded with a discussion of the interpr tation of the results in terms of the geometrical properties of sign space.

At a time when the space lag is exciting so much comment, is interesting to examine this volume with a view to the probab existance of a "communication theory lag." The framework with which Kotel'nikov solves his problems is one which we now ca statistical decision theory. Statistical decision theory was develope in this country by Wald shortly before Kotel'nikov first publishe his results in 1947. It was eight years later, however, before Middl ton and Van Meter became the first to apply Wald's ideas to cor munication problems in this country.

Two other Sputniks of note in this volume are the geometric interpretation of signals and the analysis of nonwhite noise filteri problems by the use of inverse filters. The former was first publishe in this country by Shannon in 1949, the latter by Bode and Shanno in 1950.

It seems fair to say that at the present time there is no sing major concept (with the possible exception of the material in Sectio 9–2) in Kotel'nikov's work which is not known in this countr As noted above, this was not true as little as five years ago. Th is not to say that the results have all been worked out in this cou try—they have not.

NORMAN M. ABRAMS
Dept. of Elec. Engr
Stanford Universi
Stanford, Cal

Abstracts

This Section of the issue is devoted to abstracts of material which may be of interest to PGIT members. Sources are Government, Industrial and University reports, and books and journals published outside of the United States. Readers familiar with material of this nature which is suitable for abstracting are requested to communicate the pertinent information to one of the Editors or Correspondents listed below.

omb Filters Using Delay Elements for Periodic Signal Detection—. Aoyagi and T. Kasami (in Japanese). (*J. Inst. Elec. Commun. ngrs. Japan*, vol. 43, pp. 32–38; January, 1960.)

This paper is concerned with optimum comb filters for the improvement of the SN Ratio of a repeating signal corrupted by oise. An optimum comb filter is defined as a filter consisting of given number of delay elements having a delay time equal to ne signal period, τ, and several adders, multipliers and phase verters, such that a maximum SN Ratio improvement is obtained, ubject to the stability condition.

A comb filter containing N delay elements has a transfer function the form $T(s) = T_z(z) = f(z^{-1})/g(z^{-1})$, where $z = e^{s\tau}$. The following assumptions are made: 1) the stability requirement can be expressed as a limitation on the locations, p_k, of the poles of $T(s)$, i.e., $Re\ p_k \leq -\delta = \log \alpha < 0$, where α depends on the accuracy and stability of the elements; and 2) the noise is additive and uncorrelated with the signal, and its power spectrum may be considered to be flat in intervals $2\pi n/\tau \leq \omega \leq 2\pi(n+1)/\tau$, n an integer.

It is proved that the signal-to-noise power ratio improvement factor, G, is given by $G = T_z^2(1)/$the sum of the residues of the poles of $T_z(z)T_z(z^{-1})z^{-1}$ inside the unit circle $|z| = 1$. For a given and α, a simple method to obtain the optimum transfer function and its corresponding G is found.

The modification for the case of a finite repetition of the signal discussed, and it is shown that if N is very much less than the number of repetitions of the signal, results similar to those for the ideal case hold. The last section is devoted to a discussion of the deviation from the ideal case due to inaccurate delay times, and a simple evaluation formula for the deviation is derived.

inciples of Sorting—D. A. Bell (in English). (*Computer J.*, vol. pp. 71–77; July, 1958.)

The entropy measure of disorder, and the reduction of entropy use of information, are shown to be applicable to the task of ting items from random sequence to a specified sequence. In rticular, the number of operations in sorting by merging illustes the relation between number of binary selections and rection of entropy. In sorting on electronic computers, one may h to exchange storage requirement against number of operations.

Spectrum of Television Signals—D. A. Bell and G. E. D. Swann (in English). (*Wireless Engr.*, vol. 33, pp. 253–256; November, 1956.)

Analysis of video signals from the BBC transmitter at Sutton Coldfield showed that sidebands are grouped round the harmonics of line frequency with 50-cps spacings, but since successive line harmonics differ by a multiple of 25 cps, the fringes of these groups overlap so as to give components at 25-cps spacing in the center of each interline gap. The amplitude of video component falls rather faster than as $1/f$. Data are given for the statistical distribution of instantaneous amplitudes at particular video frequencies.

The Rate of Transmission of Information in Pulse-Code Modulation Systems—A. R. Billings (in English). (*Proc. IEE*, pt. C, vol. 105, pp. 444–447; September, 1958.)

By calculating the information communicated through a noisy PCM channel (taking account of the remaining equivocation), it is deduced that the rate could be 70 per cent – 80 per cent of channel capacity, though with low error rates it is less than 33 per cent. It is suggested that error-correcting codes be devised to allow working in the condition of high equivocation of individual digits.

Communication Efficiency of Vocoders—A. R. Billings (in English). (*Electronic and Radio Engr.*, vol. 36, pp. 449–453; December, 1959.)

Spoken English of high intelligibility is taken to be equivalent to printed English and hence Shannon's work ("Prediction and Entropy of Printed English") can be used to infer that the communication rate of speech is approximately 20 bits/sec. A SN Ratio of 24 db is postulated for high intelligibility, and compared with the Shannon channel capacity, a vocoder has an efficiency of the order of 1 per cent. Efficiency is improved by introducing volume limiting so that the intensities of signals indicating the presence of various formants are independent of the formant intensities. In the simple "low-power vocoder," the speech is modulated on to a 10-kc carrier which is then limited before spectral analysis. In the "tertiary low-power vocoder," separate AGC amplifiers are provided in each of the three frequency channels which are expected to receive high-intensity formants.

Radar Range Performance—D. L. Drukey and L. C. Levitt. (Hughes Aircraft Co., Culver City, Calif., Technical Memorandum 560, August 1, 1957; revision of Technical Memorandum 277, April 15, 1954.)

This report treats the case of a pulse radar system in which several echoes are obtained from the object as the radar beam scans over it. The best method of processing the receiver output so as to distinguish signal plus noise from noise alone is determined. Various types of fading echo signals, as well as constant ones, are considered. Expressions are obtained which permit calculation of the probability of detecting the target as a function of range and the radar parameters. Several nonoptimum processing schemes are also investigated, as are schemes which make use of variable radar parameters to obtain better performance. The question of choosing radar parameters is also studied briefly. Appendixes indicate in considerable detail the derivations underlying the results presented.

Noise-Reducing Codes for Pulse-Code Modulation—J. E. Flood (in English). (*Proc. IEE*, pt. C, vol. 105, pp. 391–397; September, 1958.)

To overcome the weighting by significance of digit of the noise associated with error in a pulse in binary PCM, it is proposed to reduce the error-risk on the more significant digits by using more than one pulse per digit. The effect of noise is computed for a range from 1 to 10 pulses per digit. Alternatively, longer pulses could be used for the more significant digits. Since the bandwidth must be increased, there is more noise and no over-all improvement.

Comma-Free Codes—S. W. Golomb, B. Gordon, and L. R. Welch (in English). (*Can. J. Math.*, vol. 10, no. 2, pp. 202–209; 1958.)

This study concerns codes which can be deciphered work by word without the whole message being available. Consider an alphabet consisting of the numbers 1, 2, $\cdots$, n. A set of k-letter words is called a comma-free dictionary if whenever $(a_1a_2 \cdots a_k)$ and $(b_1b_2 \cdots b_k)$ are in the dictionary the "overlaps" $(a_2a_3 \cdots a_kb_1)$, $(a_3a_4 \cdots a_kb_1b_2)$, etc., are not in the dictionary. An upper bound is obtained for the number of words in the dictionary. It is shown that this upper bound can be attained for odd $k \leq 15$, and it is conjectured that the upper bound can be attained for arbitrary odd integers k. Some asymptotic results for large n are also obtained. For odd k and $n \to \infty$, the authors show that the number of words in the dictionary is approximately $k^{-1}n^k$ (as compared with n^k possible words when there is no restriction on overlapping).

The Effect of Random Noise Background upon the Detection of a Random Signal—H. S. Heaps (in English). (*Can. J. Phys.*, vol. 33, pp. 1–10; January, 1955.)

A noise distributed in phase and power according to a Rayleigh law is studied in terms of its effect upon the detectability of a signal of similar phase and amplitude distribution. An expression is derived for the probability distribution of the ratio of the power of the signal plus noise to that of the noise in the absence of the signal. The corresponding result is given for the ratio of the averages over several observations. Also derived is the probability distribution of the fractional change in noise plus signal power due to a given fractional change in signal power.

The Effect of a Random Noise Background upon the Detection of a Sinusoidal Signal—H. S. Heaps (in English). (*Can. J. Phys.*, vol. 33, pp. 509–520; August, 1955.)

A sinusoidal signal is examined after reception upon a noise background of random phase and power distributed statistically according to a Rayleigh law. An expression is obtained for the probability distribution of the ratio of the power of the signal plus noise at any instant to the power of the noise which would be received at that instant in the absence of the signal. The corresponding result is given for the ratio of the averages over several observations. The equations contain as a parameter the ratio of the signal power to the mean noise power. When each observed value of the power is the average over a small time interval, the formulas are applicable provided the noise and the signal have the same frequency. A similar analysis is presented to deal with the case in which the noise and the signal have different frequencies and in which each observation is the average over a small time interval. Comparison with the results of a previous paper, in which the signal was assumed to have a Rayleigh distribution in pha and power, indicates the effect of extreme fluctuation of an ori nally sinusoidal upon its resultant with a random noise backgrou

Optimum Filter Functions for the Detection of Pulsed Signa in Noise—H. S. Heaps (in English). (*Can. J. Phys.*, vol. 36, p 692–703; June 1958.)

This paper is concerned with the optimum method of processi a signal received upon a background of noise. Determination made of the transfer function of the process that maximizes t ratio of the average power of n successive samples of the outp signal to the mean output noise power. For sufficiently large valu of n, the ratio is a close approximation to the ratio of the signa to-noise energy contained in a sample of the output over a fini time. The maximum ratio is calculated when the input is a re tangular dc pulse upon white noise and when it is a cosine pul upon nonwhite noise.

The transfer function to maximize the ratio of signal volta to noise power was determined in a previous paper. It is fou that for the rectangular pulse the two methods of optimizati lead to very similar ratios of signal-to-noise energy and that t third-order low-pass Butterworth filter produces a ratio of signa to-noise energy that lies within a few per cent of the theoretic maximum. Such is not the case for the cosine pulse.

Optimum Network Functions for the Sampling of Signals in Noise H. S. Heaps and M. R. McKay (in English). (*Proc. IEE*, pt. vol. 105, pp. 438–443; September, 1958.)

In contrast to the classical "detection" principle of produci a maximum value of SN Ratio at a single instant, it is often pref able to take the average of a number of samples or to take t average over a continuous interval. Filters are specified for the taks and for inputs consisting of: 1) rectangular pulse upon wh noise, and 2) cosine pulse upon nonwhite noise.

On the Interpolation and Prediction of Signals plus Noise f Infinite and Finite Smoothing Times—D. McDonnell and R. Perkins (in English). (*Proc. IEE*, pt. C, vol. 106, pp. 47–54; Marc 1959.)

The weighting function $h_1(t)$ which represents the operation the desired filter is separated into parts, $h(t)$ which is free fr impulses and $h_2(t)$ containing all the impulses. The resulting fo of the Wiener-Hopf equation is then solved by the Laplace tra form and two examples are worked for the usual requirement minimizing the mean-square error, averaged over infinite tim Section III deals with the alternative requirement to minimize t ensemble average of the error at a specified time after the appli tion of signal to the circuit.

An Emphasis Scheme which is Information Theoretically Match to a Continuous Communication Channel—H. Miyakawa Japanese). (*J. Inst. Elec. Commun. Engrs. Japan*, vol. 42, p 1220–1226; December, 1959.)

It is well established that the apparent noise spectrum of continuous communication channel can be transformed by emphasis scheme. However, the ordinary emphasis scheme is ve inefficient from the information theoretical point of view. T paper presents a new emphasis scheme which is matched to t channel information-theoretically.

The message signal, which is assumed to be white, is sampl at Nyquist intervals and its sample values are supplied to a clas cal emphasis circuit, whose characteristics may be written as

$$S(\omega) = 1 + \sum_{n=1}^{\infty} r_n e^{-in\tau\omega}$$

where τ is the Nyquist interval. The power of the emphasiz sample value is greater than that of the original message by

$$10 \log_{10} \left(1 + \sum_{n=1}^{\infty} r_n^2\right),$$

so the emphasized sample values cannot be sent into a chan when channel power is limited. This is the reason that the classi emphasis scheme is inefficient. In the new emphasis scheme, t

mphasized sample values are further processed by a" sawtooth"-pe nonlinear circuit. The inverse nonlinear circuit is placed in e de-emphasis system of the receiver.

It is shown that the new emphasis scheme enables one to trans-rm a channel with a given noise spectrum into a channel with arbitrary noise spectrum, within the limitations of information eory, even if the channel power is limited.

gical Elements Based on a Majority Decision Principle and the mplexity of their Circuits—S. Muroga (in Japanese). (*J. Inst. ect. Commun. Engrs. Japan*, vol. 42, pp. 993–1000; November, 59.)

A logical element based on a majority decision principle is defined re as an organ such that a majority of the binary values of its put decides a binary value of its output. Theoretical aspects logical elements of this sort were discussed by McCulloch and tts in 1943 and later by von Neumann in 1959, where neurons ere models in both cases. This paper does not overlap with these pers. That is, the introduction of a concept of unequal input upling amplitudes gives a new aspect of this theory which may of importance in engineering applications. The parametron rks exactly on this principle and we may also be able to construct element of this sort with other components like diodes.

In this paper, after the description of a mathematical model r the element, with definitions of a threshold and a total coupling mber of the element, it is shown that when unequal coupling nplitudes of integral values are assigned to the inputs, the element n represent various functions, not only symmetrical but also symmetrical, according to the combination of values of coupling mplitudes. Though the number of these functions is limited for specified number of input variables, even a single element based the majority decision principle can represent a fairly complex nction. A theorem which specifies functional forms which can represented by a single element is given.

Consequently, synthesis of a switching circuit for a given function ith such elements has economical advantages because of the few ements required and the high speed of operations due to the nall number of cascaded elements required from the input to e output. The discussion of synthesis is divided into two cases, cording to the restriction of the maximum number of inputs hich are coupled to any element in the circuit. In the first case, here there is no restriction, the required number of elements is eatly reduced to construct the circuit for a given function. In e second case, where the maximum number of inputs to any ement in the circuit is specified as a certain value, for example, ve in this paper, the synthesis still requires less elements than e required number of relay contacts which was shown by Shannon. ast, it is shown that unique features of the majority decision inciple are indicated in the synthesis of a circuit for a symmetrical nction or a partially symmetrical function, requiring extremely w elements.

Statistical Study of Fading, Diversity Effects and the Improvement Characteristics of Diversity Receiving Systems—M. Nakagami Japanese). (Shukyosha Co., Kobe, Japan; 1947.)

This book summarizes the principal results of a series of the thor's statistical studies, performed during the years from 1935 1943, on fading, diversity effects and the improvement available om a diversity system. The contents are divided into two parts: perimental studies (Chapters 1 to 5) and theoretical studies hapters 6 to 11).

In Chapter 1, the necessity for a statistical consideration of ding is emphasized, and some methods of treating the character-ics of fading and diversity effects, as well as the improvement aracteristics available from diversity reception, are presented. Chapter 2 are discussed the new photometric and some other nple methods of observing fading and diversity effects statistically. me statistical properties of fading and diversity effects, observed the above method, are shown. In Chapter 3, the author describes e characteristics of space, frequency and polarization diversity ects observed in the HF band with various propagation distances, rections and frequencies. In Chapter 4, the observed coherent aracter of HF waves is presented. From these, the writer suggests e structure of the wave coming from various distance more than 00 km. In Chapter 5, the characteristics of the improvements tained in space, frequency, and polarization diversity systems are shown and compared with each other. In Chapter 6, the author presents a mathematical method of treating fading and diversity effects. In Chapter 7 are established general theories of the interference, attenuation and polarization types of fading as well as of mixed types of fading. The observed intensity distribution is compared with that theoretically derived. In Chapter 8, general theories of coherence and the space diversity effects are discussed in detail. The influence of the correlation, due to the wave picked up by feeder lines, on the magnitude of the space diversity effect observed in the input of the receivers, is discussed. In Chapter 9, similar theories of polarization diversity are developed. In Chapter 10, the writer establishes a general theory which enables one to estimate the improvement available from n-diversity combination in various systems, such as linear addition, switching, etc. In Chapter 11, the improvement characteristics of space and polarization systems used for long distance HF communication are discussed in detail. The essential points described in this book are summarized in the last chapter.

The m-Distribution as the General Formula of Intensity Distribution of Rapid Fading—M. Nakagami, K. Tanaka, and M. Kanehisa (in English). (*Memoirs of the Faculty of Engineering, Kobe University, Japan*, no. 4; March, 1957.)

This paper summarizes the principal results of a series of the authors' statistical studies in the last five years on fading characteristics, especially on the intensity distributions due to rapid fading.

The method of derivation and the important characteristics of the m distribution (a form of the gamma distribution), originally found in the author's experiments and designated by them, are described. The applicability of this formula to both ionospheric and tropospheric modes of fading is well confirmed by some experimental results. Its theoretical background is also discussed in detail. A theoretical interpretation of the log-normal distribution is given on the basis of this formula. An extremely simplified method of estimating the improvement characteristic of various systems of diversity reception is presented. The mutual dependences between the m formula and other basic distributions are fully discussed. Some generalized forms of the basic distributions and their dependences on the m form are also investigated. Two methods of approximating a given function with the m distribution are also shown.

The joint distributions of two variables following the m distribution are derived. Using this, a unified theory of diversity effects is established. These theories have enabled the authors to estimate the improvement characteristics due to diversity reception for two correlated signals having various fading ranges.

An entirely new shorthand method of observation of fading is proposed based on the theory of the m distribution. Using this apparatus, the authors are now obtaining much valuable information on fading.

An Analysis of Non-Mathematical Data Processing—E. A. Newman (in English). (In "Mechanization of Thought Processes," Her Majesty's Stationery Office, London, England, p. 865; 1959.)

A great deal of data-processing involves the recognition of pattern and the judgment whether patterns are alike. Patterns can be built up from elementary marks, but the number of arrangements which can be constructed from a given set of elements of reasonable extent is so large that matching against an exhaustive set is impracticable. It is important to use the regularities of the patterns that arise in practice, and to have readily available in store common subgroups from which the larger patterns can be constructed. The store organization and structure should be modified with time as experience shows which of the whole range of patterns are the most probable. If a machine cannot find from its store patterns to match those being presented to it, it must create new patterns to try, and to do this it must be noisy.

A New Spectrum Computer "Meriac-1-F" for the Analysis of Recorded-Curve Data—M. Nishida and T. Furuhata (in Japanese). (*J. Inst. Elec. Commun. Engrs. Japan*, vol. 42, pp. 1045–1050; November, 1959.)

In this paper the outline of the data-processing machine "Meriac-1-*F*" is described and some analytical results are shown.

This machine has been designed and constructed to reduce manual efforts and obtain calculated results speedily, and functionally consists of two sets of apparatus: the input device "Meriac-1-

F-100" and the output device "Meriac-1-*F*-200." The former is a kind of information-transforming device which automatically reads the values from a curve of complicated data on a given chart at high speed without missing any physical information and records them on a six-unit binary digital tape.

The output device is an analog computer for Fourier analysis which automatically reads the given binary digital tape at ultra-high speed and sucessively records at high speed the amplitude of the wave for each component frequency.

Modulation by Random and Pseudo-Random Sequences—R. C. Titsworth and L. R. Welch (in English). (Jet Propulsion Lab., Pasadena, Calif., Progress Rept. 20–387; June 12, 1959.)

This report deals with the modulation of signals by discrete random and random-like sequences which may change state only at integral multiples of some basic time division t_0. The signals may be modulated (sampled) in many fashions, depending mainly upon the types of sequences and signals available, the desired output phenomena, and the sequential rate.

In general, a sequence may sample a set of signals at random, or it may sample in some fixed deterministic fashion. Furthermore, deterministic processes may be constructed to possess certain random-like qualities. Special attention is given to random Markov chains and linear pseudo-random sequences; the signals selected for modulation are not restricted to any one class, and examples are given for sinusoids and square waves.

Specifically, the effects of carrier-signal waveform and type of sequence upon the over-all power spectrum are considered. In the case of sinusoidal modulation, the effect of phase shift is investigated.

Conditional Probability Computing in a Nervous System—A. M. Uttley (in English). (In "Mechanization of Thought Processes," Her Majesty's Stationery Office, London, England, p. 119; 1959.)

The author examines the hypothesis that the organization of nervous systems is based on the two principles of *classification* and *conditional probability*. Given a system of binary inputs, the facility of connecting the inputs in nearly all possible ways, and the availability of delays, it is possible to construct a network having separate indicators for each of a number of spatial and temporal patterns. This is a classification network, but it becomes a conditional-probability system if in addition, the presentation of any one pattern causes the computation of the conditional probability of every other pattern. The building up of records of conditional probabilities, or "learning," implies some mechanism of slow change in the state of a neuron. Electric-circuit models are suggested which exhibit slow recoveries analogous to the spontaneous recovery of conditioned reflexes.

On the Distribution of the Product of Diode Detector Waveforms—E. L. R. Webb (in English). (*Can. J. Phys.*, vol. 34, pp. 679–691; July, 1956.)

The probability distribution of the product of two waveforms such as come from the diode second detectors of radio receivers is examined over the whole range of SN/R's. Computed curves of probability density are given for small and moderate values of SN Ratio and the limiting form for large signal-to-noise indicated. The pure noise case is the only one immediately available in terms of tabulated functions. Compared to the Rayleigh distribution it rises much faster, reaches its maximum sooner and lower, and decays much more slowly. The very large SN Ratio case approaches án impulse function. Estimates of mean and variance are given.

A Game Theoretic Model of Communication Jamming—L. R. Welch (in English). (Jet Propulsion Lab., Pasadena, Calif., Memo. 20–155; April 4, 1958.)

The communication jamming problem is analyzed from a game-theoretic viewpoint. A model is described in which both the communicator and jammer transmit real numbers at equally spaced intervals which are taken to be the unit of time. Both parties have power limitations, and the jammer has the entire past history of the communicator's signal available for analysis.

It is shown that, if the signal power is 1 and the jamming power is J, the communicator can transmit at an information rate arbitrarily close to $\frac{1}{2}\log_2(1 + J)/J$ bits per unit time with an arbitrarily small probability of error of a message block. Furthermore, the jammer can prevent the communicator from diong any better than this.

A Class of Definitions of "Duration" (or "Uncertainty") and t Associated Uncertainty Relations—M. Zakai (in English). (Mem randum of the Scientific Department, Ministry of Defence, Isra September 14, 1959.)

A new class of definitions for "time duration" and "bandwidt (or "time uncertainty" and "frequency uncertainty"), in terms norms of L^p spaces, is suggested. Some properties of the definitio and the associated uncertainty relations are derived. As exampl of the application of these concepts, the problem of the approa of the probability distribution of shot noise towards the norm law and the "beamwidth"—"aperture width" product in anten theory are considered.

The following two volumes are collections of papers published the A. S. Popov Scientific-Technical Society for Radio Engineeri and Electrical Communication, Moscow, USSR. They were edited V. I. Siforov. The tables of contents are given in full below, but beca of space considerations, abstracts are given in only a few cases. papers are in Russian.

Issue No. 2—1958

On the Determination of General Technical Characteristics Communication Systems—A. G. Zyuko.

On the Determination of the Amount of Information $I(\mathfrak{n}, \xi)$ a Random Object $\mathfrak{n}$ concerning a Random Object ξ in the Case Continuous Transmission—G. B. Linkovsky.

On the Theory of an Ideal Receiver in the Sense of Kotelnikov B. A. Varshaver.

Basic Relations in Radio Receiving Systems Employing Integrati and Filtration of Signals in the Presence of Fluctuation Noise N. L. Teplov.

On the Theory of Radiocommunication Channels with Multipa Propagation—V. I. Siforov.

Estimation of the Largest Possible Value of the Entropy of Unknown Distribution Characterized by Several Moments—B. Fleishman and G. B. Linkovsky.

Estimation of the Entropy and Distribution Function of a Sca Random Variable Characterized by Several Sample Moments G. B. Linkovsky.

Several Questions in the Theory of Construction of Error Correcti Codes—L. F. Borodin.

Several general methods for the construction of error-correcti codes are evolved through the use of the concepts and techniq of number theory. Specifically, methods for the detection a correction of errors are developed, and a single error-correcti code is considered in detail. For the case where the number levels, a, is a prime number, a systematic procedure is given the construction of a number of special codes (*e.g.*, optimal cc for an erasure channel; a code in which all sequences differ fr one another in exactly $d = (a - 1)a^{m-1}$ positions; a single-err correcting code). Finally, two circuits for adders mod a are analyz

Accumulation of Disturbances and Fading in Main Radio Re Lines—V. I. Siforov.

Cross-Talk Noise Arising in FM Radio Relay Lines Due to Multipa Propagation or Mismatch and Non-Uniformity in Antenna Feeders A. V. Prosin.

On the Analysis of Multichannel Communication Systems w FM and Bandwidth Compression—A. V. Prosin.

Approximate Analysis and Modelling of Accumulation of D turbances in Radio Relay Communication Links—Yu. B. Sindl **Parametric Methods in Electrodynamics**—L. A. Druzhkin.

On the Question of Choice of the Form of the Vector-Paramet Equation in the Solution of Problems of Distribution of Cha on Closed, Cylindrical, Infinite Conductors, and on Linear, Pla Closed Conductors—L. A. Druzhkin.

Issue No. 3—1959

I. Theory of Communication Channels with Randomly Varying Characteristics.

ptimal Reception of a Parameter Transmitted over a Channel th Additive, Multiplicative and Phase Disturbances—V. I. forov, B. S. Fleishman, and G. B. Linkovsky.

Optimal reception of a parameter transmitted over a channel th additive Gaussian noise was first considered by V. A. otelnikov for the case of a single parameter, and by D. Slepian for e case of many parameters. In recent years, the growing interest in ansmission of metereological and, particularly, geophysical param-ers, frequently under conditions which have not been explored fficiently (for example, in connection with satellites and high-alti-de rockets), suggests the need for the development of a theory of ise immunity for more general types of channels.

It appears that a natural generalization of a continuous channel provided by the recent work of V. I. Siforov on channels with ndomly varying characteristics. Such channels are found in nnection with tropospheric and ionospheric propagation at UHF well as in ordinary short-wave propagation. In all these cases, ultipath propagation of radio waves takes place, with attendant ndom modulation of transmitted signals both in amplitude and ase in the presence of internal noises in the receiver. This gives e to channel noise which has multiplicative, additive and time-lay components. Previous investigations of channels of this type re information-theoretic in nature and were concerned with e estimation of their capacities. Present work is concerned with ise immunity in the case of optimal reception of a parameter rough such a channel, treated as a problem of optimal estimation a parameter in the sense of mathematical statistics.

n the Ideal Reception of a Parameter Transmitted over a Channel mprising a Small Number of Paths—G. B. Linkovsky.

n the Problem of Optimal Statistical Estimation of the Character-ics of a Multipath Communication Channel—B. S. Fleishman, B. Linkovsky, and Yu. B. Sindler.

II. Theory of Information

n a Method of Linear Coding with Error Correction of Trans-itted Signals—R. R. Varshamov.

n the Comparison of Uniform Codes for Binary Transmission—A. Varshaver.

onstruction of an Optimal Code in the Sense of Shannon in the mplest Case of a Binary Noisy Channel—B. S. Fleishman.

In this work, an optimal code in the sense of Shannon is con-ructed for a binary symmetric channel. The construction is complished by random selection of M input sequence of length with subsequent elimination of some of them. Cases where such mination can be avoided are considered. Estimates of the prob-ility of obtaining an optimal code by a random selection as well the probability of correct decoding for finite n are obtained. is shown that in all cases, the former probability tends to unity th increase in n much faster than the latter probability.

perimental Study of Statistical Characteristics of Patterns—N. Osher.

III. Theory of Radio Relay Communication Laws.

n the Influence of the Form of Correlation Function of Inhomo-neous Turbulence in the Troposphere on Scatter Propagation UHF—A. V. Prosin.

vestigation of the Properties of Probability Distributions of ding and Disturbances in Radio Relay Communication Lines—. B. Sindler.

Device for Measuring the Correlation Coefficient in Long Line mmunication at UHF—I. P. Levshin, and G. I. Slobodenuk.

The following papers were published singly by the Professional oup on Information Theory (I) and the Professional Group on tomata and Automatic Control (A) of the Institute of Electrical Communication Engineers of Japan, 2-8, Fujimicho, Chiyodaku, Tokyo, Japan. All are in Japanese, but English abstracts are given below when available. The affiliation of each author is given so that interested readers may contact the author directly for further information.

Basic Theory for Pattern Recognition (A; December 10, 1959)—T. Iizima (The Electrotechnical Laboratory, 1, 2-chome, Nagata-cho, Chiyuda-ku, Tokyo.)

Theory of Tape-Sorting (I; March 18, 1960)—T. Iizima (See above.)

Theory of Sequential Machines (A; February 18, 1960)—S. Fujino (School of Science, Kyushu University, Hakozaki-machi, Fukuoka, Kyushu.)

A Theory of Waveform Prediction (I; January 19, 1960)—T. Kasami (Engineering Department, Osaka University, 9-chome, Higashinoda-ku, Miyakojima-ku, Osaka.)

The problem of selecting out the signal wave from a finite time observation is discussed. The prediction operator, which gives the predicted waveform based on the observed waveform, is assumed to be one of finite dimension which is generally realizable. The cost of observation and prediction is partly considered. A minimax solution is obtained for the case when the first and second statistics and a quadratic loss function are assumed.

Information-Theoretical Analysis of Classification (I; January 19, 1960)—Z. Kiyasu and S. Ikeno (Electrical Communication Lab., 1551 Kichijoji, Musashino-shi, Tokyo.)

Improvement Efficiency of an Error-Correction System in Short-Wave Circuits (I; February 26, 1960)—T. Kumagaya, K. Teramura, and H. Sakaguchi (Japanese Overseas Radio and Cable System, 1–5, Ohte-machi, Chiyoda-ku, Tokyo.)

Proof of Mathematical Theorems Using a Computing Machine (A; March 31, 1960)—N. Kuroda (Nagoya University, Chigusa-ku, Nagoya.)

Basic Planning of a Vocal Typewriter (A; January 18, 1960)—K. Maeda, et al. (Kyoto University, Honcho Yoshida, Sakyo-ku, Kyoto.)

Error Probability in FS Radio Teletype System (I; February 26, 1960)—S. Miyake, K. Tadenuma, and T. Nakai (Japanese Overseas Radio and Cable System, see above.)

Sampling Errors in the Measurement of Autocorrelation (I; March 18, 1960)—H. Miyakawa (Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo.)

Not only the mean and variance, but also the statistical properties of γ_n as a function of n are calculated, where γ_n is an estimate of autocorrelation $\rho(n)$. An estimate is derived from a single sample, of finite length, of the time series which is assumed to be stationary and Gaussian. It is found that the spectral decomposition theorem can be applied to the difference $\gamma_n - \rho(n)$, though the difference is a nonstationary time series as a function of n. A simple relation between the spectral density of $\gamma_n - \rho(n)$ and that of the original time series is also derived.

A Psychological Study for the Improvement of Teletypewriters (I; December 19, 1959)—G. Ohwaki and K. Maruyama (Psychological Laboratory, Tohoku University, Sakura-koji, Sendai.)

Programming for Proof of Mathematical Theorems (A; March 31, 1960)—G. Shimanouchi (Tokyo University of Education, 24 Ohtsuka, Kubomachi, Bunkyo-ku, Tokyo.)

Binary Representation of Vowels (I; January 19, 1960)—H. Suzuki and M. Oh-izumi (Electrical Communication Laboratory, Tohoku University, Sakura-koji, Sendai.)

Programming for the M-1 Computer for the Proof of Mathematical Theorems (A; March 31, 1960)—T. Takasi (Electrical Communication Laboratory, 1551 Kichijoji, Musashino-shi, Tokyo.)

An Investigation of Radar Tracking Error (I; January 19, 1960)—K. Tanaka (Engineering Department, Kobe University, 1-chome, Mizukasa-cho, Nagata-ku, Kobe.)

NOTICE TO ADVERTISERS

IRE TRANSACTIONS ON INFORMATION THEORY will accept advertising. For full details contact E. K. Gannett, The Institute of Radio Engineers, Inc., 1 East 79 Street, New York 21, N. Y.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT TRANSACTIONS. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the PROCEEDINGS OF THE IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE TRANSACTIONS ON INFORMATION THEORY is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates.

All technical manuscripts and editorial correspondence should be addressed to Arthur Kohlenberg, Melpar, Inc., 11 Galen Street, Watertown 72, Mass. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to David Van Meter, Litton Industries, Inc., Waltham, Mass.

INSTITUTIONAL LISTINGS

The IRE Professional Group on Information Theory is grateful for the assistance given by the firms listed below and invites application for Institutional Listing from other firms interested in the field of Information Theory.

IBM RESEARCH, INTERNATIONAL BUSINESS MACHINES CORP., Yorktown Heights, N. Y.
Error Correcting & Detecting Codes, Theory of Assemblies & Automata, Information Networks, Reliability

REPUBLIC AVIATION CORP., Farmingdale, N. Y.
Aircraft, Missiles, Drones, Electronic Analyzers; U. S. Distr. of Alouette Turbine-Powered Helicopter